GENERAL LIBRARY
AUG - 6 1920
UNIV. OF MICH.

Journal of Educational Research

Official Organ of the National Association of Directors of Educational Research

B. R. BUCKINGHAM, *Editor*

Associate Editors

E. J. Ashbaugh — W. W. Charters — George D. Strayer
Leonard P. Ayres — J. A. Courtis — Lewis M. Terman
Walter S. Monroe

VOL. I APRIL, 1920 No. 4

CONTENTS

Constructing a Language and Grammar Scale, *W. W. Charters*..... 249
Psychological Tests and the Probable School Success of High-School Pupils, *W. M. Proctor*..... 258
Retardation in One-Room Rural Schools in Kansas, *F. J. Kelly* and *A. K. Loomis*..... 271
A Group Scale of Intelligence for Use in the First Three Grades, *Luella W. Pressey*..... 285
The Application to Tables of Distribution of a Shorter Method for Computing Coefficients of Correlation, *L. P. Ayres*..... 295
Editorials..... 299
Reviews and Abstracts..... 302
Bibliographies..... 311
News Items and Communications..... 321
National Association of Directors of Educational Research.... 325

Application has been made for entry as second class matter at the Post Office at Menasha, Wisconsin

Journal of Educational Research

Copyright 1920, by
Public School Publishing Co.

Published for the Bureau of Educational Research, University of Illinois

Editor-in-Chief: **B. R. Buckingham**

Director of Bureau of Educational Research, University of Illinois

Associate Editors

R. J. ASHBAUGH, *Ex-Officio, Secretary of Research Association, University of Iowa, Iowa City*
LEONARD P. AYRES,
Ed'l Sec'y, Russell Sage Foundation, N. Y. C.
W. W. CHARTERS,
Carnegie Inst. Tech., Pittsburgh
S. A. COURTIS,
City Schools, Detroit, Mich.
WALTER S. MONROE,
University of Illinois, Urbana, Ill.
GEORGE D. STRAYER,
Columbia University, New York City
LEWIS M. TERMAN,
Leland Stanford Junior University, Stanford University, Calif.

Contributing Editors:

VIRGIL E. DICKSON,
Director of Research, Public Schools, Oakland, California.
CHARLES FORDYCE,
Dean Teachers' College, and Director of Bureau of Educational Measurements, University of Nebraska.
WILLIAM S. GRAY,
Dean College of Education, University of Chicago.
M. E. HAGGERTY,
Director Bureau of Co-operative Research, University of Minnesota.
V. A. C. HENMON,
Director School of Education, University of Wisconsin.
ARTHUR W. KALLOM,
Assistant Director Department of Educational Investigation and Measurement, Public Schools, Boston, Mass.
F. J. KELLY,
Dean, School of Education and Director Bureau of School Service, University of Kansas.
ROBT. H. LANE,
Assistant Superintendent of Schools, Los Angeles, Calif.
WILLIAM A. McCALL,
Assistant Professor of Education, Teachers' College, Columbia University.
GEORGE W. MELCHER,
Director Bureau of Research and Efficiency, Public Schools, Kansas City, Missouri.
EUGENE A. NIFENECKER,
Assistant Director Bureau of Reference, Research, and Statistics, Public Schools, New York City.
JOSEPH P. O'HERN,
Assistant Superintendent of Schools and Director of Efficiency Bureau, Public Schools, Rochester, New York.
LEON O. SMITH,
Assistant Superintendent (in charge of Educational Research) Public Schools, Omaha, Nebraska.
CLARENCE W. SUTTON,
Director Division of Reference and Research, Public Schools, Cleveland, Ohio.
W. W. THEISEN,
Supervisor of Educational Measurements, State Department of Public Instruction, Madison, Wisconsin.
CLIFFORD WOODY,
Assistant Professor of Education, University of Washington.

THE JOURNAL OF EDUCATIONAL RESEARCH is published at 450 Ahnaip St., Menasha, Wis. monthly except July and August by Geo. Banta, Official Printer to the Public School Publishing Co. The subscription price is $3.00; $3.20 in Canada or Mexico; and $3.40 in other foreign countries.

Application has been made for entry as second-class matter at the post-office at Menasha, Wis.

The editorial office is at Urbana, Illinois. Address all communications on educational matters, articles for the magazine, etc. to B. R. Buckingham, University of Illinois, Urbana, Illinois.

Articles to Appear Soon

Intelligence Tests for College Entrants by *E. L. Thorndike.*

The Operation of Bureaus of Educational Research, by *W. W. Theisen.*

Reproduction as a Measure of Reading Ability, by *A. W. Kallom.*

Use of Psychological Tests in the Guidance of High-School Pupils, by *W. M. Proctor.*

Putting Standardized Tests to Practical Use in Rural Schools (a series of articles), by *S. S. Brooks.*

Educational Measurements in a Small System, by *F. H. Koos.*

Journal of Educational Research

Editorial Office, Urbana, Illinois — Editor-in-chief B. R. Buckingham, University of Illinois

VOLUME I — APRIL, 1920 — NUMBER 4

CONSTRUCTING A LANGUAGE AND GRAMMAR SCALE

W. W. CHARTERS
Carnegie Institute of Technology

INTRODUCTION

In our studies of the efficiency of school children in using correct speech, two difficulties were recognized at almost the beginning of the investigation. In the correction of language forms upon a language level without knowledge of grammatical rules and definitions, it was found that few teachers had definite information about the errors which their children made. They recognized that errors occurred and that in a general way some were more common than others; but any detailed knowledge of exactly how such errors should be attacked was almost entirely lacking. This lack of definite information prohibited them from making local attacks upon specific errors; and this is, in almost all cases, with some marked exceptions, true not only of the individual teacher but also of whole school systems. For, the tendency is to take pot shots at errors and this in spite of the fact that the only method of correction is to attack them one at a time and that a certain number of errors should be assigned to each grade for frontal attack with the supporting assistance of the grades above and below.

It was found also, that the study of grammar seemed to have an almost negligible effect upon the correction of errors. A steady increase of skill from grade to grade is noticeable before grammar is studied but in the grades where grammar is taught and where a rapid increase in skill is expected there did not seem to be a greater increase than in the lower grades in which instruc-

tion is not given. If this is true, then two results must follow. Either grammar should be eliminated as a grade subject of study or methods should be used by which it will have a noticeable effect upon language usage. Very serious doubts as to its usefulness in grades are growing rapidly. But until it is clearly demonstrated that it does not and cannot affect the correction of errors, the doubt will not be resolved.

The probable reason for its failure to function as a basis of correction is that the educational world has taken the indefensible position that incorrect forms must not be placed before children. That the verb must agree with its subject in person and number is taught in connection with the analysis of correct sentences only. For in select educational circles it is considered bad form to place on the board sentences such as, "Was you looking for something?" or "He don't look well" and ask children to apply the rule to their correction. This is probably caused by a confusion of thought. It is entirely defensible to argue that the incorrect form should not be made conspicuous when memory alone is involved. Before the study of the grammatical rule which is found broken the pupil may become confused by seeing both the correct and incorrect form given equal prominence; but once the rule has been taught and memory is reinforced by reason, there is every reason for fusing the error and the rule into one experience. Only by this process, carried on in the school room, will the pupil recognize that rules have any use as correctives. Until that is done deliberately and thoroughly, grammar will be concerned merely with the analysis of correct sentences, and errors will live in a world apart. And if this is not done, grammar must disappear from the grades. It is difficult at its easiest, and because it is difficult, it leads to confusion and probably, as mental exercise, does more harm than good. If it can demonstrate its utility as a corrective, it may hope to live. Like other vested interests it must demonstrate its ability to render service to the public.

Following the realization of the two facts that language errors are not attacked methodically and persistently and that grammar does not seem to function, the idea was developed that formal tests might be devised which would place information upon both these points at the disposal of the teacher. Such tests should obviously be diagnostic and should display the ability of the

pupils to recognize incorrect forms, to correct them and to give the reason for correction where this depended upon rules of grammar already learned.

METHOD

By language tests are meant those in which the pupil is expected merely to recognize the correct form and by grammar tests are meant those in which he can provide in addition, the basic rule for correction couched in grammatical terminology. To say that "He gave it to me and Mary" is incorrect because it is "impolite" or does not "sound right" is not a use of grammar. The grammatical rule states that the first personal pronoun stands last in a series; for only in this latter case is the terminology of grammar used.

In constructing the language tests several alternatives presented themselves. The first form considered consisted of the use of alternative forms, the incorrect one of which was to be crossed out, as "*(Was) (Were) you looking for something?*" This form could be quickly treated by the pupils and rapidly scored by the teacher. But it would not give an accurate display of the normal ability of the pupils to use the correct form. For normally, when an individual is about to ask the question, "*(Was) (Were) you looking for something?*" he does not have the two forms in his field of attention; if they were both there the hint would be sufficient to supplement his memory in the choice of the correct one.

This form was discarded and the device accepted of stating, "In the following sentences some are correct and some are incorrect. On the line beneath the sentence write the correct form." It was felt that the pupil would thereby have to make a complete and unassisted decision in a situation approximating very closely to the conditions of normal speech.

In the grammar test the same sentences are used but in addition space is provided on the right half of the test page for giving the reasons for any corrections made. The form is shown in the following portion of one of the tests. The complete test has forty items. The title and the provisions for general information about the class and the pupil are omitted.

GENERAL INSTRUCTIONS

After telling the children not to turn the papers over, distribute the folders, this side up. Have each pupil fill in the blanks at the top. Then say: "Look at the directions while I read them to you."

DIRECTIONS: (To be read by the teacher and pupils together)

"This test is given to pupils who have studied language and grammar to see how well they are able to tell when sentences are right and when they are wrong. Now look at the sample below."

1. I told him to go. ....

"The plan is to read this sentence over carefully and see if it is right. If it is right make a cross on the dotted line below the sentence. The sentence, 'I told him to go,' is right so we shall make a cross on the dotted line below it. Make the cross now." (The teacher should pause to allow time for doing this and should pass around the room in the lower grades to make certain every pupil understands.)

"If the sentence is not right we are to put the words on the dotted line below it. Let us try one that is not right."

2. I told she to go. ....

"The right form is 'I told her to go.' So we shall write 'her' on the line below it."

"On the right-hand side we are to tell why we make the change on the left-hand side. We always begin our reason for the change with the word *Because*. In this case we write—*Because 'her' is in the objective case, object of the verb, told*. (The teacher should pause until the pupils write the reason at the right of the sentence.) We must not write too large and we may use all of the lines on the right-hand side so that we may have plenty of room for giving our reasons."

"We shall now try a third exercise."

3. It is hisn. ....

"The right form is *It is his*. We make the change *Because his is the possessive form of the pronoun, he*. Fill this in correctly, writing 'his' on the line below and beginning the reason on the right-hand side with the word *Because*. Do this now."

"Remember that if the sentence on the left is all right, we do not need to write anything on the right-hand side because we have to make no change and so have to give no reason for making a change."

The teacher should pass around the room and see that the children fill these in correctly and when all have finished he should say, "Begin work on the exercises below. Stop when you have finished this page and bring your paper to me." This is not a speed test so ample time should be allowed to each to finish. This test should be completed in two periods during the same day. Part I should occupy the first period, and Part II the second. At the

beginning of the second period, go over the above instructions so far as may be necessary before directing the children to begin on Part II. After the children have finished Part II, collect the papers.

1. John went to town.	1 .....
2. May Inez and me go?	2 .....
3. It teaches a person something you may use.	3 .....
4. Yous girls play too hard.	4 .....
5. When one lives in town they hear noises.	5 .....
6. That is hern.	6 .....
7. Who do you want?	7 .....
8. Him and I will do it.	8 .....
9. Us boys did it.	9 .....
10. The boy was raised by its own mother.	10 .....
11. These books are yourn.	11 .....
12. It was a book with riddles in them.	12 .....
13. Are those them?	13 .....

Content of the Tests

From studies made in Kansas City, Detroit, and elsewhere, a rather complete list of common grammatical errors in the verba-

tim expressions of the children was obtained. These were used and selection was made upon the basis of frequency of occurrence. The following tests were constructed in duplicate forms.

I. Language	II. Language and Grammar
Pronouns	Pronouns
Verbs A	Verbs A
Verbs B	.....
Miscellaneous	Miscellaneous

Each test has forty items, one or two of which are correct, to keep faith with the children in conforming to the promise that they would find that some of them were correct. The number, forty, was determined partly by the size of practicable pages and partly by the fact that it gives a comprehensive list of the errors at hand. The forty do not include all items as do the tests of the fundamental operations in arithmetic, but all frequently occurring and many infrequently noticed errors are included so that for purposes of class instruction, where only the commoner ones will be studied, as distinguished from individual errors made by only one or two, the significant ones are undoubtedly included. Some pupils will make other errors but all of sufficient frequency to receive class attention are present in the tests.

For purposes of comparison of efficiency at different times two forms of the test (Form I and Form II) are available. The forms are alike in that the same types of errors are included in each and in the same order but differ in the sentences in which the errors are found.

The language tests are constructed so as to be given from grades III to VIII or in the high school. The grammar tests will naturally be given only in those grades which studied the grammatical facts involved.

Time Limit

In giving the tests no time limit is prescribed at present. This has been omitted because the sentences necessarily vary in length if approximately natural sentences are used; and relative ability, as shown by the test, would be greatly affected by speed in handwriting. In the grammar division of the test time conditions are further disturbed by the fact that no standardized form for stating grammatical rules has been accepted or proposed by the

schools. In addition to these considerations opposing the setting of a time limit is the further fact that the test being diagnostic in character it is essential that all the errors should be attacked by each pupil.

However, in our most recent modification of the language section of the tests we are experimenting with the writing of corrections by asking the pupils not to write the *whole sentence* in correct form but to insert, on the line beneath, the correct *words*.[1] This may make it possible to set a time limit on the language section; but it does not remove any difficulty in the grammar section.

Weighting

A very careful study was made of approximately three thousand one hundred seventh and eighth-grade test papers on pronouns to determine the weight of each item. The plan followed was to determine the number and percent of incorrect replies and using sigma (σ) as a basis and 2.5σ as zero to give the approximate weight to each. Fortunately, it was found that the relative ranking of individuals was not appreciably affected by dropping the two decimals from the weighted amount and using the nearest integer. Furthermore, it was discovered that the correlation between rankings determined upon the bases of unequal and equal weights for each item was slightly over 0.90. Consequently, it was decided to adopt the simple plan of counting each item as possessing unitary value.

It is entirely possible that if zero ability were more closely defined than 2.5σ and if it were based on zero third-grade ability rather than upon zero seventh-grade skill, as was done in our study of the three thousand one hundred cases, the weighting might be more significant than has been found in the latter case. But for diagnostic purposes, which is the professed function of these tests, this consideration is not of great significance.

What was true of the weighting of language skill when based on percentages of failure was more noticeable in grammar weightings because of the very large percentages of failure to give reason for correction. The three thousand one hundred papers examined had a median percentage incorrect of over fifty.

[1] The directions reproduced on page 252 above are the ones used in this modification of the original test.

Scoring

The scoring of the language section presents no serious difficulty since the forms tested are those upon which any well-informed teacher is competent to judge. However, scoring cards are provided to expedite the work.

In the grammar section the absence of standardized forms, as in addition or multiplication, presents a rather cumbersome difficulty. For instance, the sentence "Him and I will do it" was written correctly as "He and I will do it." The reasons given for changing to the correct form, *He*, were the following among many.

(1) Because *He* is in the nominative case, subject of the verb *will do*.
(2) Because *He* is the subject of the sentence.
(3) Because *He* is in the nominative form.
(4) Because *He* is subjunctive form.
(5) Because it would not be right to say *Him will do it*.
(6) Because we should use *He* instead of *him*.
(7) Because *He* is the noun.
(8) Because it seems to be turned around.
(9) Because *me* is objective form, object of the verb *and*.

It is obvious that (7), (8), and (9) are absurd and that (5) and (6) give no grammatical basis. It is also clear that (1) is a complete answer consisting of both the description of the case (nominative) and the syntactical relation governing it (subject of the verb "will do"). The difficulties of the decision of all such cases is seen in (2), (3), and (4). This difficulty arises from the fact that they each include one of the two parts of the complete answer. The second (2) gives the syntactical relation; (3) gives the case form and (4) is, probably, a mixture of (2) and (3).

It was decided that while teachers should naturally expect a pupil to give both elements as a standard of class instructions, an examiner should rate the answer as correct if one or the other of the two elements were given. This would give a more accurate guage of ability than would the elimination of (2), (3), and (4). For there is no standard form for the answer and if the pupil knows the one element he probably knows the second since they are interdependent, and in any case either one of the two elements affords a good clue for practical speech.

The norms to be established for these scales will include not only the median but, in addition, the 25 and 75 percentile. This is necessary in any diagnostic scale. For since it focuses attention on the individual, it is obvious that no stimulation will be derived from its use by the half of the class above the median if only the median is used as a standard.

The Use of the Tests

These tests are designed to serve four purposes. In the first place, a teacher who gives one of them to his class and has them marked and tabulated has a clear idea of the points at which to work in the acquiring of language skill, an idea much clearer than that afforded by incidental and more or less casual observation. In the second place, it is possible by the use of the alternative forms to see what progress is made during any period. In the third place, they provide a means of discovering whether or not grammar is of any use in the control of errors, as can be easily found by testing freshman students in high school who have and have not studied grammar in the grades. Finally it presents ocular proof to individual pupils that they do not know how to use certain forms with accuracy and ease.

These tests are still in tentative form and the cooperation of those interested is invited for assistance in smoothing out the defects and in establishing standards for the grades and the high school. Copies of the tests and other information may be obtained from the Bureau of Educational Research at the University of Illinois, Urbana, Illinois.

PSYCHOLOGICAL TESTS AS A MEANS OF MEASURING THE PROBABLE SCHOOL SUCCESS OF HIGH-SCHOOL PUPILS

W. M. PROCTOR
Leland Stanford Junior University

The validity of the Stanford-Binet Scale, when applied to high-school pupils, has already been discussed by the writer in the issues of *School and Society* appearing October 19, and 26, 1918.[1] In those articles it was shown that very significant correlations had been obtained between intelligence quotients,[2] (I. Q.'s), resulting from the individual tests of 107 high-school pupils and the school marks earned by the same pupils during the school year, 1916-1917; also between I. Q.'s and teachers' estimates of intelligence made during the same year.

Two years and a half later there were 66 of the original 107 high-school pupils remaining. Teachers who had known all of these pupils during their stay in the high school were asked to give estimates of their intelligence upon the same rating sheet as that which was used in 1916-1917. All school marks earned during the two and one-half years were averaged. Correlations were then found (a) between the I. Q.'s obtained in 1916-1917 and the teachers' estimates made in 1919; (b) between the average of all school marks earned up to April 1, 1919 and I. Q.'s obtained in 1916-1917; and (c) between the average school marks and the teachers' estimates made in 1919. Table I shows the close agreement between the correlations obtained in 1916-1917 and those found in 1919.

Table I shows that the correlations obtained in 1918-1919, when the same comparisons were made as in 1916-1917, were

[1] Proctor, W. M. "The use of intelligence tests in the educational guidance of high-school pupils," *School and Society*, 8:473-78, 502-9, October, 19-26, 1918.

[2] The intelligence quotient is obtained by dividing the mental age by the chronological age. Thus a twelve-year-old, chronologically who tested eight years old mentally would have an I. Q. of 0.66, expressed for convenience "66." The I. Q. is an index of relative brightness. For further discussion of intelligence quotient see: Terman, L. M. *The intelligence of school children*, Houghton Mifflin and Co., 1919, p. 7.

TABLE I. COMPARISON OF CORRELATION OBTAINED IN 1916-1917 AND THOSE OBTAINED IN 1918-1919[a]

Year	Correlation between I. Q. and Teacher Estimates	Correlation between I. Q. and School Marks	Correlation between School Marks and Teacher Estimates	Total Number of Cases
1	2	3	4	5
1916-17.....	0.586±0.043	0.545±0.046	0.702±0.033	107
1918-19.....	0.583±0.055	0.487±0.063	0.667±0.046	66

[a] Pearson's formula[2] (shorter method) was used in making all correlations.

practically as high as those obtained in the first instance. The results of two and one-half years of follow-up work seem to indicate that the person who made the original tests with the Stanford-Binet Scale in 1916-1917, would have been in a position to give very helpful advice to all of the pupils tested by him; also that his predictions as to the possible educational future of each of these pupils would have deserved serious consideration by parents and teachers.

As a means of discovering individual differences between school children in order that they may be grouped in classes according to ability, the individual psychological test has been shown to be a helpful tool. From the standpoint of school administration, however, the individual test presents serious difficulties. The time required to give an individual test to a high-school pupil varies from 40 minutes to 120 minutes. The total number of pupils that can be examined by a single examiner in a day will seldom exceed ten. The use of the Stanford-Binet abbreviated scale enables an examiner to test from 15 to 25 pupils in a day. Even so, it is impossible to use the individual method when a rapid survey of an entire school population is to be undertaken.

Group mental examinations afford the only means of meeting the demand for a speedy and reliable method of measuring the mental abilities of large groups of people. Under the supervision

[2] Rugg, H. O. *Statistical methods applied to education*. New York: Houghton Mifflin and Co., 1917, p. 274.

of Dr. L. M. Terman the writer directed the giving of Examination *a* (Form A) and Examination Alpha (Form 5) of the Army Scale to 1,349 high-school pupils, representing eight California high schools, during the school year 1917-1918.

Examination *a* consists of ten tests: (1) Oral Directions, (2) Memory for Digits, (3) Disarranged Sentences, (4) Arithmetical Reasoning, (5) Information, (6) Synonym-Antonym, (7) Common Sense, (8) Number Series Completion, (9) Analogies, (10) Number Comparison. The total possible raw score is 237. This test was given to portions of the first-year high-school classes in the Oakland Technical, Oakland Central and Fremont, Oakland, high schools and to all of the pupils present on the day of the examination at the Palo Alto Union High School. The total number of pupils was 715.

Group Examination Alpha consists of eight tests: (1) Oral Directions, (2) Arithmetical Reasoning, (3) Practical Judgment, (4) Synonym-Antonym, (5) Disarranged Sentences, (6) Number Series Completion, (7) Analogies, (8) Information. This test was given to all pupils present on the day of examination in the San Mateo, Redwood City, Mountain View, and Santa Clara union high schools. The total number of these pupils was 624.

These group mental examinations were applied to all of the pupils in each of the high schools enumerated above at exactly the same time. A sufficient number of examiners, trained by Dr. Terman, were taken to each high school, to cover the entire high school in one forty-five minute period. The size of the groups ranged from 40 to 150. It took the writer and his assistant a total of 134 hours to test 107 high-school pupils by the individual method. Six trained examiners were able to give Examination *a* to 350 Palo Alto high-school pupils in 45 minutes. The test blanks were scored by university students. Their work was carefully checked and the results tabulated by the writer.

I. Group Test Results Compared with Individual Test Results

One hundred and sixteen of the high-school pupils tested with Examination *a* had previously been given the Stanford-Binet Scale. Table II makes comparison of the two kinds of mental examination. Although no Binet I. Q.'s are found in the group 140-149, six Army Scale I. Q.'s are between 140 and 149. This

is because a higher mental age is attainable on the Army Scale than on the Binet Scale. For example, a high-school boy fifteen years and two months old passed every test in the Stanford-Binet

TABLE II. RELATION BETWEEN THE I. Q.'S OF 116 HIGH-SCHOOL PUPILS EARNED ON THE STANFORD-BINET SCALE AND THE I. Q.'S OF THE SAME PUPILS EARNED ON EXAMINATION *a*, ARMY SCALE

I. Q.'s on Stanford-Binet Scale	I. Q.'s on Examination *a*, Army Scale							Totals
	80–89	90–99	100–109	110–119 (*Median*)	120–129	130–139	140–149	
1	2	3	4	5	6	7	8	9
130–139					3	3	5	11
120–129			1	3	7	9	1	21
110–119 (*Median*)		2	8	11	9	2		32
100–109			13	3	4			20
90–99		7	11	8				26
80–89		5						5
70–79	1							1
Totals	1	14	33	25	23	14	6	116

Median for Binet I. Q.: group, 110-119
Median for Army Scale I. Q.: group, 110-119
Correlation, Pearson's formula, 0.736; P. E., 0.029

Scale, thus earning a mental age of nineteen years and six months, and an I. Q. of 129. On the Army Scale, Examination *a*, he made a raw score of 219, corresponding to a mental age score of twenty-one years and eleven months, and an I. Q. of 144.[4] In other words, the Stanford-Binet Scale does not give the superior high-school pupil an opportunity to earn as high an I. Q. as he can earn on the Army Scale. This factor would tend to lower the

[4] Mental age norms for both tests of the Army Scale were worked out by Dr. Samuel Kohs and the writer. It was found that about fifteen points on Examination *a* and twelve points of raw score on Examination Alpha corresponded roughly to a mental age year. Possible raw score, Examination *a*, 237, possible mental age twenty-three years; possible raw score, Alpha 212, possible mental age twenty-four years and nine months.

correlation between the two sets of I. Q.'s. The correlation obtained in Table II (+0.736) is a strong indication that if the Stanford-Binet Scale is a valid means of finding the mental level of high-school pupils, the Army Scale Examination *a* is also valid for the same purpose.

II. Army Scale Results Compared with High-School Marks

The school work of all the high-school pupils examined by means of the two army group tests was carefully followed up for the school years 1917-1918 and 1918-1919. The marks given were reduced to a comparable basis by assigning arbitrary values to each type of mark employed by the various high schools. Thus an "A" or a "1" was given a value of 95; a "B" or a "2," a value of 85, etc. Letters or numbers with plus and minus signs were given intermediate values. All of the marks earned by a given pupil were averaged, but no case was included in the tables unless the marks for at least two semesters of school work, the equivalent of one year, were available.

1. *Army Scale, Group Examination* a.—Table III makes a comparison between the I. Q.'s obtained from the Group Examination *a* and the quality of high-school work of 494 high-school pupils. The total number taking the test was 715, but only 494 cases had ratings for one year of school work.

Since the correlation obtained (+0.343) is 12.8 times the indicated P. E., it has considerable significance. While it is not as high as the correlation between Binet I. Q.'s and school marks found in Table I, there are several factors which may have tended to lower the correlation. The army tests were designed for use with soldiers in cantonments. Many of the questions have to do with matters of common knowledge about a military camp, but with which high-school pupils have no acquaintance. This would especially affect the scores of high-school girls. There are twenty cases falling in I. Q. groups below 95, where the indicated school work is of a quality of 80 percent or above, and fifteen of these cases, or 75 percent, are girls.

Another factor which may have tended to lower the correlation is the skewness of the school marks curve toward the high percents. The median school marks group in Table III is 80-84.

TABLE III. CORRELATION BETWEEN THE I. Q.'S OF ARMY GROUP EXAMINATION *a* AND THE QUALITY OF SCHOOL WORK OF 494 HIGH-SCHOOL PUPILS

School Marks	Army Group Examination *a* I. Q.'s										Totals
	84 or Lower	85–89	90–94	95–99	100–104	105–109 (*Median*)	110–114	115–119	120–124	125 or Above	
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
90 or above	.....	.....	3	2	2	3	6	6	6	6	34
85–89	.....	.....	5	9	18	24	22	24	8	4	114
80–84 (*Median*)	1	2	9	28	41	46	30	13	9	3	182
75–79	1	3	9	19	19	19	19	7	2	.....	98
70–74	2	3	4	14	11	11	7	1	1	.....	54
65–69	.....	1	1	4	1	1	2	1	.....	.....	11
55–64	.....	.....	.....	.....	.....	1	.....	.....	.....	.....	.....
Totals	4	9	31	76	92	105	86	52	26	13	494

Medians: I. Q.'s group, 105-109; school marks, group 80-84
Semi-interquartile ranges: I. Q.'s, 6 points; school marks, 4½ percent
Correlation, Pearson's formula, 0.343; P. E., 0.027

This represents a grade of "B" or "2." It is necessary for these high-school pupils to receive marks of "B" or "2" in all subjects required for university recommendation. The Palo Alto high school is in close proximity to Stanford University, and the Oakland high schools are in the immediate vicinity of the University of California. The fact that 68.8 percent of the grades given to these 494 high-school pupils were "B" or above is an indication that teachers were influenced in their marking by the demand for "B" grades for university recommendation. Many pupils with just *average* ability were given marks ranking *superior*. Hence there would be a rather wide difference between their mental ability as shown by the tests and their school progress as shown by their marks.

2. *Army Scale, Group Examination Alpha.*—Table IV compares the I. Q.'s of 480 of the high-school pupils of San Mateo,

Redwood City, Mountain View, and Santa Clara with their school marks. There were 624 pupils belonging to these high schools who took Examination Alpha, but only 480 cases came under the rule requiring ratings for one whole year of school work.

TABLE IV. CORRELATION BETWEEN THE I. Q.'S OF THE ARMY GROUP EXAMINATION ALPHA AND THE QUALITY OF SCHOOL WORK OF 480 HIGH-SCHOOL PUPILS

School Marks	Army Group Examination Alpha I. Q's.										Totals
	84 or Lower	85–89	90–94	95–99	100–104	105–109 (*Median*)	110–114	115–119	120–124	125 or Above	
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
90 or over	.....	.....	.....	3	3	15	12	9	9	5	56
85–89	.....	.....	.....	8	17	15	24	13	6	6	89
80–84	.....	.....	4	6	22	21	20	10	5	1	89
75–79 (*Median*)	.....	.....	7	25	33	23	10	7	4	.....	109
70–74	.....	4	10	18	14	22	12	1	1	.....	82
65–69	1	3	3	12	7	8	8	1	.....	.....	43
55–64	.....	.....	2	5	3	1	1	.....	.....	.....	12
Totals	1	7	26	77	99	105	87	41	25	12	480

Medians: I. Q.'s group 105-109; school marks, groups 75-79
Semi-interquartile ranges: I. Q.'s, 6 points; school marks, 6½ percent
Correlation, Pearson's formula, 0.413; P. E., 0.026

The correlation obtained in Table IV (+0.413) is 15.9 times the indicated P. E. and 0.07 higher than the correlation found in Table III. The higher correlation found in this table may be due to the fact that Examination Alpha was the result of careful revision of the first series of tests in the light of preliminary experimentation in three army cantonments and the returns from the tests of several thousand school children. Also, the high schools in which Examination Alpha was applied were not in such close proximity to universities as were the high schools represented in Table III. The group median for school marks is 75-79, and only 48.8 percent of the ratings given in these four high schools ranged as high as "B" or over.

When account is taken of the possible reasons for the difference in the correlations as shown in Tables III and IV, it is safe to say that they are of approximately equal value. For use in the public schools Examination Alpha is the better scale, because it consists of but eight tests, can be given in less time, is scored more rapidly, and costs less to print.

III. Individual and Group Mental Tests as Means of Indicating Probable Retention or Elimination of High-School Pupils

1. *Elimination among pupils tested with the Stanford-Binet Scale.*—On the first day of April, 1919, it was found that 41 of the original 107 pupils tested in 1916-1917 with the Stanford-Binet Scale had dropped out of school, leaving 66 still in attendance. Table V gives the distribution of the 107 cases, showing the I. Q.'s of pupils who left high school to go to work, of pupils who were transferred to other high schools, and of pupils who still remain in the same high school.

TABLE V. DISTRIBUTION ON APRIL 1, 1919, OF 107 HIGH-SCHOOL PUPILS TESTED WITH STANFORD-BINET SCALE IN 1916–1917

Binet I. Q.'s Earned in 1916–1917	Total Number of Cases	Distribution April 1, 1919, by Percents		
		Out at Work (%)	Out, Transfer'd to Other High Schools (%)	Remaining in Same High School (%)
1	2	3	4	5
79 or lower	1	100	0	0
80–89	7	72	14	14
90–99	29	31	9	60
100–109	27	22	22	56
110–119	22	0	27	73
120–129	15	0	13	87
130 or above	6	0	33	67
No. of cases	107	21	20	66
Median I. Q.'s		94	110	110

The only pupil testing below 80 I. Q. dropped out at the end of the first semester of 1916-1917 to go to work. All of those in group 80-89 who left school to go to work, did so by the end of the first year. Failure in school work has been recognized as the most fruitful cause of elimination from high school, but the relation between mental ability and failure in school work has not heretofore been given due consideration. At the end of two and one-half years none of those testing below 80, and only 28 percent of those testing 80-89 remain in high school. On the other hand, 100 percent of those testing 110 or over are pursuing their studies either in the Palo Alto high school or in other high schools. When the average school ratings of the different groups is taken into account the close connection between mentality and elimination will be still more apparent. The average school rating of the 21 who left school to go to work was 73 percent; of the 20 transferred to other high schools, 77 percent; and of the 66 remaining in the Palo Alto high school, 79 percent.

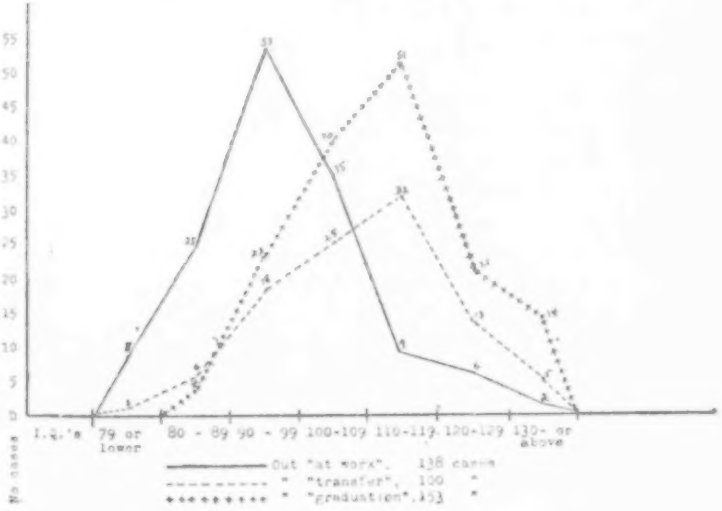
2. *Elimination among pupils tested with the Army Examination a and Alpha.*—Only five of the eight high schools where the Army Scale was applied were selected for follow-up work in connection with elimination because the records of the three Oakland high schools did not indicate whether the pupil leaving had been transferred to another high school or had gone to work. In the case of the Palo Alto, San Mateo, Redwood City, Mountain View, and Santa Clara high schools, it was comparatively easy to secure reliable data with reference to every pupil who took the test. The principals of all these schools had served in their respective positions from four to fourteen years, knew their pupils thoroughly, and had on record information concerning the movements of those who had left school since the giving of the army tests in 1917–1918.

Table VI gives for those tested with the Army Scale a distribution similar to that contained in Table V for the 107 tested with the Stanford-Binet Scale. The 955 pupils of the five high schools above mentioned have been distributed by I. Q.'s into four groups: (1) those who left high school to go to work, (2) those who transferred to some other high school, (3) those who graduated, (4) those remaining in high school on April 1, 1919. The follow-up work covered one and one-half years of school work. Figure 1 illustrates graphically the data of Table VI.

TABLE VI. DISTRIBUTION OF 955 HIGH-SCHOOL PUPILS ON APRIL 1, 1919, WHO WERE TESTED WITH ARMY SCALE TESTS IN 1917–1918 BY I. Q. GROUPS

Army Scale I. Q.'s Earned 1917–1918	Total Number of Cases	Distribution on April 1, 1919, by Percents			
		Out at Work (%)	Out, Transferred to Other High School (%)	Out by Graduation (%)	Remaining in Same High School (%)
1	2	3	4	5	6
79 or lower	13	61.5	7.7	0.0	30.8
80–89	73	34.3	8.2	5.5	52.0
90–99	202	26.2	8.9	11.4	53.5
100–109	283	12.3	8.8	14.1	64.8
110–119	221	4.0	14.5	23.0	58.5
120–129	101	6.9	11.8	19.8	61.4
130 or over	62	1.6	9.8	24.6	64.0
No. of Cases	955	138	100	153	546
Median I. Q.'s		96	110	111	107
Percent I. Q.'s below 100		62.4	25.0	17.6	26.5

FIGURE 1. ILLUSTRATING TABLE VI



In this table there appears to be a strong confirmation of the findings of Table V, as the following comparison will show:

TABLE VIA. MEDIAN I. Q.'S BY GROUPS

	Out at Work	Out, Transferred to Other High School	Remaining in Same High School
1	2	3	4
Table V	94	110	110
Table VI	96	110	107

TABLE VIB. PERCENT OF PUPILS IN EACH GROUP WHOSE I Q.'S WERE BELOW 100

	Out at Work	Out, Transferred to other High Schools	Remaining in Same High School
1	2	3	4
Table V	71.4	20.0	27.0
Table VI	62.4	25.0	26.5

The agreement between the two tables is the more striking when it is remembered that the 107 high-schools pupils of Table V were all first-year pupils when the tests were given, while the 955 pupils of Table VI comprise all the classes of five different schools.

The tendency, noted in the discussion of Table II, for girls to make lower scores on the Army Scale than boys accounts for the 5.5 percent of pupils with I. Q.'s of 80-89 who appear in the "out by graduation" group. These four pupils were all girls, and their average rating in school work was 83 percent.

Table VII indicates how school marks tend to correspond to mental level as indicated by I. Q.'s earned in the Army Scale tests.

Considering the "at work" cases, it appears that only the three highest I. Q. groups were doing a passing grade of work. Lack of mental ability was perhaps the most potent cause of elimination, but it was not the only cause operating in these cases. When interest in the subjects offered, application, and ambition are lacking, high scores in the mental tests are no

TABLE VII. AVERAGE SCHOOL MARKS OF 955 HIGH-SCHOOL PUPILS, DISTRIBUTED ACCORDING TO THE CAUSES OF LEAVING

Cause of Leaving	79 I. Q. or Lower	80–89 I. Q.	90–99 I. Q.	100–109 I. Q.	110–119 I. Q.	120–129 I. Q.	130 I. Q. or Higher	General Average
1	2	3	4	5	6	7	8	9
At work	65	64	72	66	70	75	87	71.3
Transferred	59	67	76	72	82	85	86	75.3
Graduated	.....	83	79	84	85	84	92	84.5
Average marks by I. Q. groups	62.0	71.3	75.6	74.0	79.0	81.3	88.3	77.3

necessary guarantee of school success. Turning now to the pupils transferred to other high schools, it appears that they made better average marks under every I. Q. group, except the very lowest, than did the "at work" pupils; but it is evident that some of them transferred to other high schools because of failure in school work. Later checking up will no doubt find them out of school entirely. All of the cases in the "out by graduation" group show high average school marks, even those whose I. Q. is between 80 and 89. These are the same four cases discussed in connection with Table VI. They illustrate how necessary it is for those employing mental tests to be conservative in accepting the results of any single test as final.

When the school marks earned by all of these "out" groups are averaged a series of marks is obtained, ascending gradually from the lowest to the highest I. Q. ratings, which indicates a definite tendency for the quality of school work to correspond to the mental level indicated by the Army Scale tests.

A further following up of the 153 who graduated from the five high schools between September, 1917 and April, 1919, brought out the fact that 94 of them were continuing their education in college, university, or normal school. The median I. Q. of those going on to higher educational institutions was found to be 116. One high-school girl, who earned an I. Q. of 140 on the Army Scale, graduated from the San Mateo high school at the age of fourteen years and five months. Because of the minimum age

limit of fifteen years for entrance to the university, she was compelled to wait seven months before continuing her education. Had her case been included in the "at college" group, the median I. Q. would have been 118.

Taking into account all of the cases of high-school pupils tested either by the individual or group method of mental examination we find the following ascending scale of median I. Q.'s as a further indication that the psychological tests disclosed the *approximate* mental level of the cases discussed in the foregoing tables: first-year high-school pupils, median I. Q., 105; high-school graduates, 111; those going on to college, 116.

Summary

1. Individual and group mental tests of the types described have been shown to be sufficiently reliable to justify their use as aids in determining the mental level of high-school pupils.

2. Group tests, such as Army Scale Examinations *a* and Alpha, make possible a preliminary mental survey of an entire high-school population at the beginning of a school year. The resulting raw scores will be found to be of great value in grouping the pupils according to ability. Such tests should always be supplemented with every other possible means of discovering the mental level. The results should be considered tentative and subject to revision in the light of later developments.

3. The high-school principal who makes such a preliminary mental survey of his pupils can be reasonably sure that 50 percent of those who test below normal will be eliminated within the first two years; that 25 percent additional of the subnormal group will have been transferred to other high schools because of failure in their school work; and that a negligible number will ever graduate. With this information at hand he can plan the curriculums of his pupils more intelligently. Discovering at the outset that from 15 to 30 percent of his pupils are incapable of succeeding in the conventional high-school subjects, he will undertake to make new adjustments to meet the situation. There will be fewer failures; more pupils will remain to take work that is adapted to their needs and capacities; and the high school will be less open to the charge of catering only to the intellectual aristocracy among its pupils.

RETARDATION IN ONE-ROOM RURAL SCHOOLS IN KANSAS

F. J. KELLY, *University of Kansas*
AND
A. K. LOOMIS, *Principal, Sumner County High School, Wellington, Kansas*

In 1918 Kansas still had 147,889 children enrolled in one-room rural schools out of a total of 405,319 children enrolled in all public schools. Of those in one-room schools about 72 percent were enrolled in schools having terms of seven months and almost all of the remaining 28 percent in schools having terms of eight months. The course of study in all these one-room rural schools is made out on a nine-year basis; and, if the children keep up with their classes, they are therefore still required to spend nine years to complete the course. Practically all schools in the towns and cities, on the other hand, have terms of nine months. The study here reported was one of a group of studies undertaken in the hope of reducing this inequality of opportunity.

In February, 1917 with the cooperation of the state superintendent and the county superintendents concerned, pupil's cards containing the following questions were sent to the teachers of all one-room rural schools in fourteen typical counties in Kansas.

PUPIL'S CARD

1. Are you a boy or a girl?.....
2. What is your name?.....
3. How old were you at your last birthday?.....
4. In what month were you born?.....
5. In what year were you born?.....
6. In what grade are you now?.....
7. Do you take all your studies in the same grade?.....
8. Have you ever attended any other school?.....
9. Does your father own the farm on which you live?.....
10. Do you intend to go to high school?.....
11. If you are a boy, do you intend to be a farmer?.....

A sufficient supply of the cards was sent so that each teacher had one for each pupil in her school.

Along with these cards, the following teacher's card was sent:

TEACHER'S CARD

1. County.....
2. District No.....
3. Teacher's name.....
4. Teacher's address.....
5. How many pupils are NOW on roll?.....
   (Do not give enrollment from first of the year)
6. Have you examined every pupil's card in regard to the three points mentioned in the directions for the teacher?.....
7. How many months of school does this district have this year?.....
8. How many months of school did this district have last year?.....
9. How far is it from this school house to the nearest high school?.....
10. How many years have you taught this school before this year?.....
11. What is your monthly salary?.....
12. State exactly what kind of certificate you hold.....
13. How many years of teaching experience have you had before this year?.....
14. Are you a high-school graduate.....
15. If not a high-school graduate, how many weeks have you attended high school?.....
16. Have you attended a state normal school? .....
17. If so, how many weeks?.....
18. Have you attended a college or university?.....
19. If so, how many weeks?.....

(OVER)

On the reverse side of the teacher's card the following instructions were printed:

TEACHER'S CARD (*Reverse Side*)

To the teachers of one-teacher rural schools:

With the approval of State Supt. W. D. Ross and Mr. J. A. Shoemaker, one of the Rural School Supervisors, I am undertaking to make a study of certain conditions in the one-teacher rural schools of Kansas. Your county superintendent is heartily in favor of the plan and I am certain that you will be willing to let the children fill out the enclosed cards at your first opportunity. The information that I have asked for on the teacher's card is very important. The teacher's card is printed on the other side of this sheet.

If you will assist in this study, please notice the following directions carefully before letting the children fill out the cards.

DIRECTIONS FOR THE TEACHERS

Let the children fill out the cards if they are old enough to do it correctly, otherwise please fill them out yourself. If you haven't received enough cards for all pupils use the cards for the older pupils, but please report the sex, name, age, and grade of every pupil for whom you send in no card.

Examine the cards after the children have filled them out to make certain of the following three points:

1. That every question has been answered.
2. That the answer to question three agrees with the answers to questions four and five.
3. That question six is answered correctly.

N. B. IN ANSWERING QUESTION SIX, COUNT A PUPIL WHO TAKES HIS WORK IN MORE THAN ONE GRADE IN THAT GRADE IN WHICH HE TAKES MOST OF HIS WORK.

Please fill out your own card carefully, answering every question. Return the cards to me in the enclosed stamped envelope within a week from the time you receive them if possible.

Very truly yours,

A. K. Loomis

1110 Vermont St.
Lawrence, Kansas.

(OVER)

It seems sure from these detailed instructions that the data given by the children were thus rendered reasonably accurate.

Of the fourteen counties, eleven sent in returns in time to be tabulated for this study. Of all the teachers of one-room rural schools in these eleven counties 54 percent sent in reports of all their children. The range of schools responding was from 30 percent in one county to 81 percent in another. The total number of pupil's cards received was 10,298, of which 5,267 were boys' cards and 5,031 were girls' cards. All the tabulations from these cards were made by Mr. Loomis and not by the teachers concerned.

Among the questions which the study undertakes to answer are:

1. What is the percentage of seven-month one-room schools and eight-month one-room schools by counties? (See Fig. 1.)
2. What is the amount of retardation in these one-room rural schools by counties, by ages, by grades, and by sexes? (See Figs. 2, 4, and 6.)
3. How does this retardation compare with retardation in city schools in Kansas? (See Figs. 3, 5, and 7, and Table II.)
4. How does the retardation in seven-month schools compare with the retardation in eight-month schools? (See Fig. 8.)
5. What is the cost of this retardation in terms of ambition for higher education? (See Fig. 9.)
6. What percentage of the children have spent their school lives in one school only, and what effect has the moving from one school to another had upon retardation? (See Figs. 10 and 11.)

7. What is the money cost of retardation in one-room rural schools? (See concluding paragraph.)

8. What is the money cost of the nine-year course of the one-room rural school as compared with the eight-year course of the city school? (See concluding paragraph.)

The answers to the above questions are largely contained in the tables and figures given on the following pages.

TABLE I. THE NUMBER AND PERCENT OF SCHOOLS REPORTING, AND THE TOTAL CARDS RECEIVED BY COUNTIES

	County											
	A	B	C	D	E	F	G	H	K	L	M	Total
No. of one-room rural schools	78	125	91	98	91	113	89	68	72	106	108	1,039
No. of schools reporting	63	82	37	59	27	66	47	43	43	51	39	557
Percent of schools reporting	81	66	41	60	30	58	53	63	60	48	36	54
No. of boys' cards rec'd	710	809	377	624	241	651	359	390	379	396	331	5,267
No. of girls' cards rec'd	667	742	400	690	223	608	302	362	375	347	315	5,031
Total no. of cards rec'd	1,377	1,551	777	1,314	464	1,259	661	752	754	743	646	10,298

FIGURE 1. PERCENT OF PUPILS WHO ARE IN EIGHT-MONTH SCHOOLS

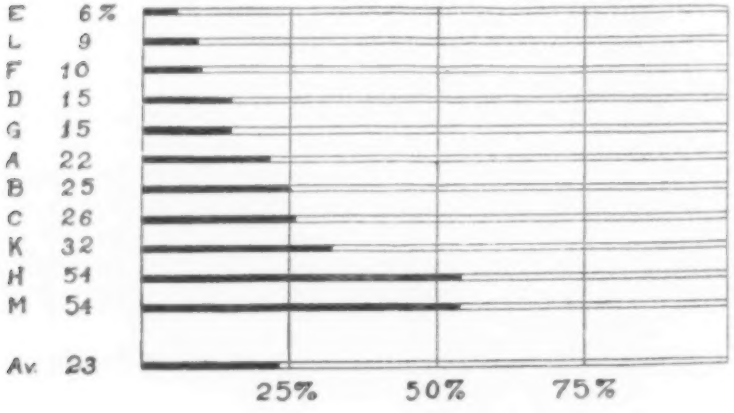


Figure 1 should be read: County E has 6 percent of its children for whom cards were received in eight-month schools, while County H has 54 percent in eight-month schools. The average of the eleven counties is also shown. Twenty-eight percent of the one-teacher rural schools of the whole state have an eight-month term. Practically all others have seven-month terms.

FIGURE 2. RETARDATION AND ACCELERATION IN 54 PERCENT OF THE ONE-TEACHER RURAL SCHOOLS IN ELEVEN COUNTIES IN KANSAS, ENROLLING 10,298 CHILDREN

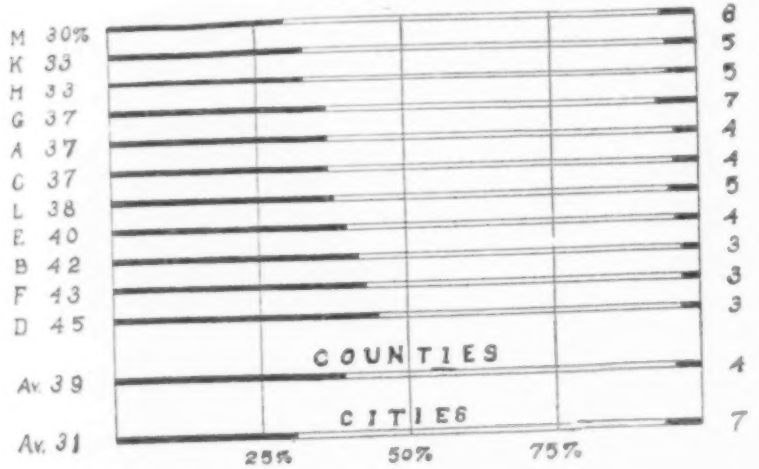


Figure 2 should be read: County M has a retardation of 30 percent and an acceleration of 6 percent. The average of the eleven counties is shown, also the average for twelve cities in Kansas in June, 1916. Normal age in grade I is six or seven at the end of school; grade II, seven or eight; grade III, eight or nine.

FIGURE 3. RETARDATION AND ACCELERATION IN TWELVE CITIES IN KANSAS, ENROLLING 28,443 CHILDREN

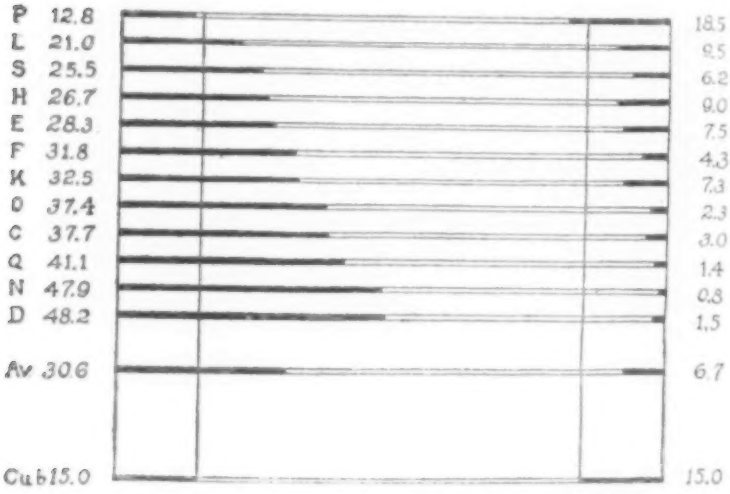


Figure 3 should be read: City P has a retardation of 12.8 percent and an acceleration of 18.5 percent. The average of the twelve cities is shown, also Cubberley's estimate of the condition under a properly adjusted course of study. The high school is included in all except N and O. Normal age in grade I is 6 or 7; grade II, 7 or 8; grade III, 8 or 9, etc.

FIGURES 4A AND 4B. RETARDATION AT EACH AGE FROM EIGHT TO FIFTEEN IN THE ONE-TEACHER RURAL SCHOOLS IN ELEVEN COUNTIES IN KANSAS

Figure 4a. Boys

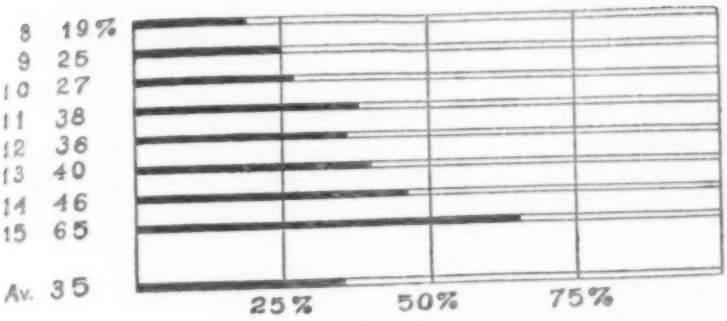


Figure 4b. Girls

Figures 4a and 4b should be read: 24 percent of eight-year-old boys are retarded and 19 percent of eight-year-old girls. The average for these ages is also shown.

FIGURES 5A AND 5B. RETARDATION AT EACH AGE FROM EIGHT TO EIGHTEEN IN TEN CITIES IN KANSAS, ENROLLING 26,608 CHILDREN

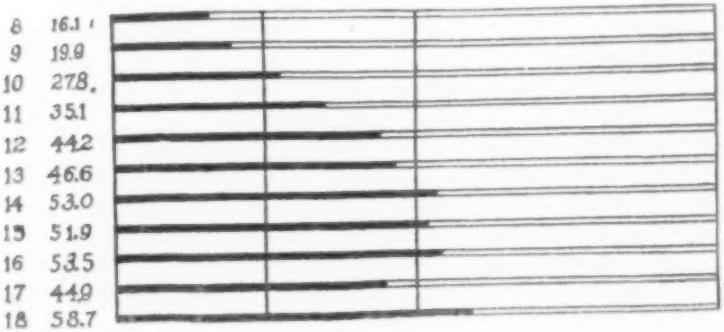



Figure 5a. Boys

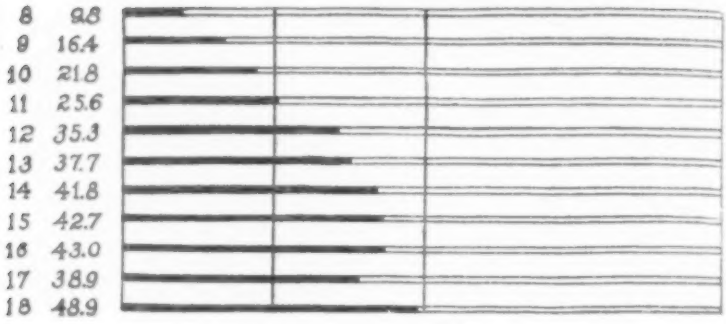


Figure 5b. Girls

Read Figures 5a and 5b as follows: 16.1 percent of eight-year-old boys are retarded, 9.8 percent of eight-year-old girls are retarded.

FIGURES 6A AND 6B. RETARDATION IN THE DIFFERENT GRADES OF THE ONE-TEACHER RURAL SCHOOLS IN ELEVEN COUNTIES IN KANSAS

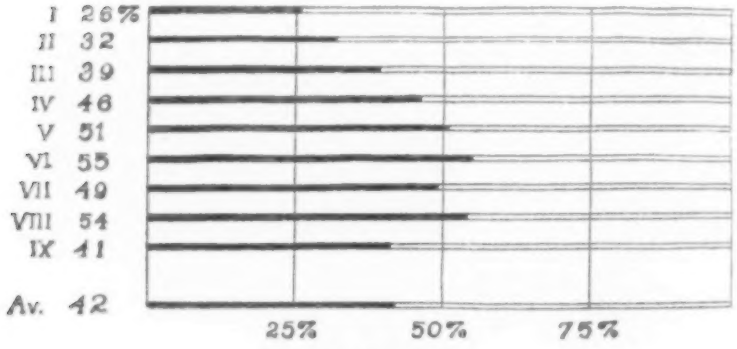


Figure 6a. Boys

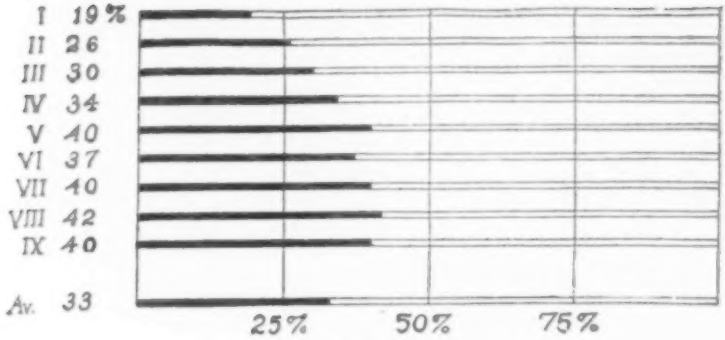


Figure 6b. Girls

Figures 6a and 6b should be read: In grade III 39 percent of the boys are retarded and 30 percent of the girls. The average for all grades is also shown.

FIGURES 7A AND 7B. RETARDATION IN THE DIFFERENT GRADES OF THE ELEMENTARY SCHOOLS IN TWELVE CITIES IN KANSAS

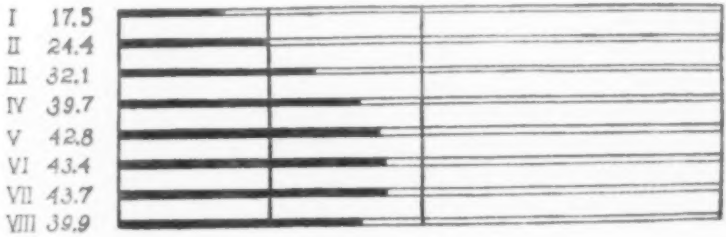


Figure 7a. Boys

Figure 7b. Girls

Figures 7a and 7b should be read: In grade III 32.1 percent of the boys are retarded, and 23.4 percent of the girls are retarded.

TABLE II. COMPARISON OF ONE-ROOM RURAL SCHOOLS AND CITY SCHOOLS IN RESPECT TO THE PROPORTION OF PUPILS RETARDED ONE YEAR AND MORE THAN ONE YEAR

	One-Room Rural Schools	City Schools
Percentage retarded one year		
Boys	28.7	18.0
Girls	25.3	16.6
Total	27.1	17.7
Percentage retarded more than one year		
Boys	17.1	14.1
Girls	10.0	10.6
Total	13.6	12.9

FIGURE 8. RETARDATION IN THE SEVEN-MONTH SCHOOLS COMPARED WITH RETARDATION IN THE EIGHT-MONTH SCHOOLS

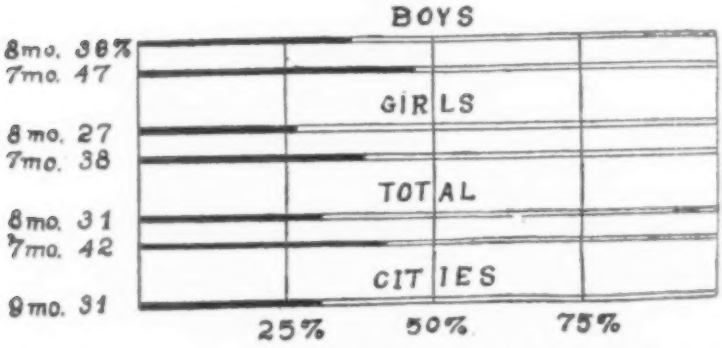


Figure 8 should be read: 36 percent of boys attending eight-month schools are retarded, while 47 percent of boys attending seven-month schools are retarded.

FIGURE 9. PERCENT OF NORMAL AND ACCELERATED CHILDREN WHO DO NOT INTEND TO GO TO HIGH SCHOOL COMPARED WITH PERCENT OF RETARDED CHILDREN WHO DO NOT INTEND TO GO TO HIGH SCHOOL

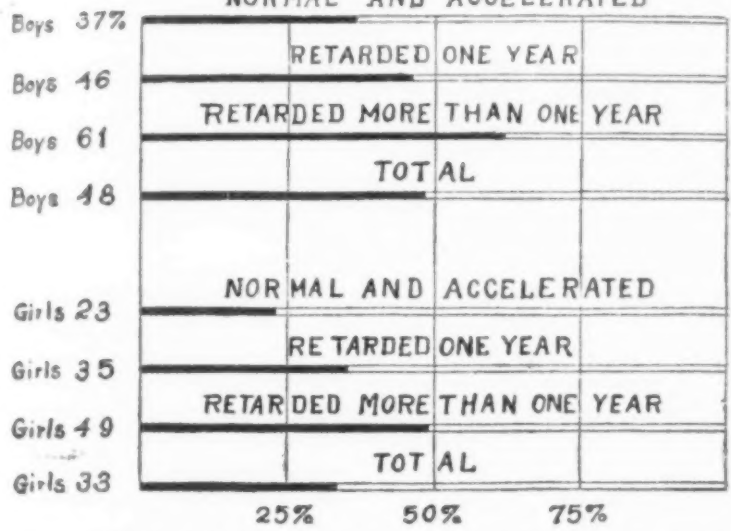
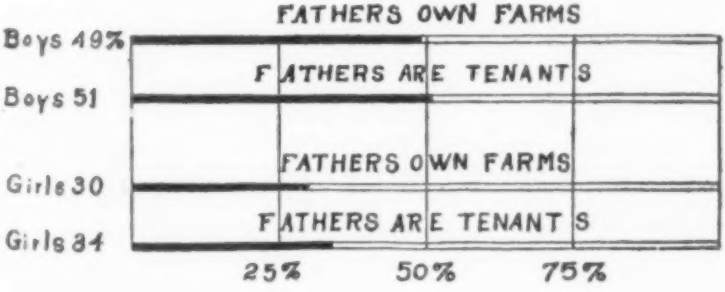


Figure 9 should be read: 37 percent of normal and accelerated boys do not intend to go to high school, 46 percent of boys who are retarded one year do not intend to go to high school, etc. Included in this comparison are 3,990 children twelve years old or older.

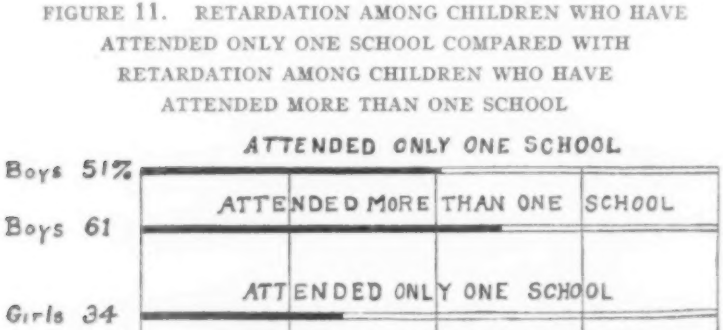
FIGURE 10. RETARDATION AMONG THE CHILDREN OF LANDOWNERS COMPARED WITH RETARDATION AMONG THE CHILDREN OF TENANTS



All children twelve or thirteen years of age who have attended only one school are included in this comparison. Figure 10 should be read: 49 percent of twelve and thirteen-year-old boys whose fathers own farms are retarded, and 51 percent of those whose fathers are tenants are retarded.

The comparison is based on children who have attended only one school in order to exclude from consideration at this point the effect of changing from one school to another. Twelve and thirteen-year-old children were used to avoid a fallacy resulting from using children of all ages. It is evident that the average age of all children who have attended only one school and whose fathers are tenants is less than the average age of all children who have attended only one school and whose fathers own farms. This follows from the fact that among children of tenants the proportion who have attended but one school decreases with the age of the children much more rapidly than it does among children of farm owners. Since retardation increases rapidly with age, the children of tenants fallaciously seem less retarded than the children of landowners when children of all ages are included.

FIGURE 11. RETARDATION AMONG CHILDREN WHO HAVE ATTENDED ONLY ONE SCHOOL COMPARED WITH RETARDATION AMONG CHILDREN WHO HAVE ATTENDED MORE THAN ONE SCHOOL



All children twelve or thirteen years of age whose fathers are tenants are included in this comparison. Figure 11 should be read: 51 percent of twelve and thirteen-year-old boys who have attended only one school are retarded, and 61 percent of those who have attended more than one school are retarded.

The Money Cost

There remains the task of calculating the costs of retardation and of conducting seven and eight-month schools for nine years. Attention is first called to the significance of a few of the main points involved.

Let it be understood that in figuring retardation exactly the same blank was used in city schools and country schools, so that the data are entirely comparable. We wish to repeat, however, that country-school children are considered of normal age for grade, if they complete the elementary-school course before they have become sixteen, while city children are regarded as one year retarded when they complete the elementary-school course at fifteen. It is clear, therefore, that country-school children from the first grade through the ninth have progressively a larger part of a year the advantage of city children before being classified as retarded. In spite of this fact, it is noted that country children have 8 percent greater retardation as a whole than city children. Not only is that true but this difference is much greater than 8 percent among the upper ages. For example, while 51.9 percent of fifteen-year-old city-school boys are retarded, 71 percent of fifteen-year-old country-school boys are retarded, and while 42.7 percent of fifteen-year-old city-school girls are retarded, 65 percent of fifteen-year-old country-school girls are retarded. It should be noted, however, that the difference is less than 8 percent when the ninth grade of the rural school is compared with the eighth grade of the city school. This indicates that the elimination is very much greater in the country school than in the city school.

Attention is also directed to the fact that while 31 percent of all children in eight-month schools are retarded, 42 percent of all children in seven-month schools are retarded, a difference of 11 percent. This means that of the present enrollment in one-room rural schools in Kansas, practically thirteen thousand are retarded in seven-month schools who would not be retarded if they were in eight-month schools. Inasmuch as only 37 percent of boys of normal age for grade of twelve years of age or above do not intend to go to high school, whereas 46 percent of those retarded for one year and 61 percent of those retarded more than one year do not intend to go to high school, the relationship between

retardation and ambition to go to high school is direct. We recognize, of course, that many children retarded are not mentally fitted for high-school work and that retardation in their cases is not the cause of their lack of ambition. But in the case of the 11 percent who are retarded in seven-month schools, who would not be retarded if they were in eight-month schools, it may be confidently stated that retardation is the definite cause of the absence of ambition to go to high school in a considerable proportion of their cases.

Assuming that a boy or girl of thirteen or more years is worth $150 a year as a producer, neglecting the loss as producers resulting from longer attendance of children each year between six and fourteen, and using thirteen-year-old pupils as a basis, the annual loss because of retardation in the one-room rural schools of Kansas is above $1,800,000. One-fourth of this amount, or $450,000, is due to the excessive retardation produced by seven-month schools above what would occur if these schools were increased to eight months. The total cost of increasing every seven-month course in Kansas to eight months would be approximately $450,000. It is clear, therefore, that the actual saving of time in retardation alone would pay the cost of making every seven-month school into an eight-month school.

Again considering the value of a child thirteen years old and older at $150 per year, the loss incurred in having country-school pupils required to attend nine years to complete the same course of study which would be completed in eight years (if the terms were lengthened from seven or eight months to nine months as in the city schools) can be estimated at $2,700,000 per year. It would cost about $1,000,000 a year to increase every seven and eight-month school to a nine-month school. There would be a saving of $1,700,000 on this score.

Naturally a very much larger proportion of the children would complete the elementary-school course if it were only eight years in length rather than nine, thus securing very much larger advantages of education than are now possible to country children. Kansas is therefore paying very dearly indeed in money, in educational opportunity, and in ambition, by persisting in maintaining country schools of seven and eight-month terms.

A GROUP SCALE OF INTELLIGENCE FOR USE IN THE FIRST THREE GRADES: ITS VALIDITY AND RELIABILITY

LUELLA WINIFRED PRESSEY
Indiana University

I. THE STATISTICAL PROBLEMS PRESENTED BY THE SCALE

In a recent number[1] of the *Journal of Educational Psychology*, the writer has presented a general descriptive account of the purpose, nature, and development of the "Primer Scale." At the time of writing that paper, little statistical evidence had been gathered as to the "validity" and "reliability" of the scale. By "validity" the writer means the extent to which the scale measures what it is supposed to measure—i.e., general intelligence; by "reliability" is meant the consistency or accuracy of the measures obtained. Thus, a scale might give very consistent measures that were not valid—that did not measure the qualities the scale was designed to measure—or, a scale might give reasonably valid measures of some quality, but the measures might be so inconsistent as to make the scale largely unsatisfactory. In other words, these two characteristics of a scale might exist in varying combinations with each other and should be kept separate in considering the value of the results. It is the purpose of the present paper to present the evidence that has been found, thus far, bearing on these two important points of (1) the validity and (2) the reliability of the Primer Scale.

II. THE VALIDITY OF THE PRIMER SCALE

1. *Sources of data.*—For a general description of the four tests composing the scale, the reader is referred to the above mentioned article. It may be briefly said that the first test is one of form discrimination in which dots are arranged in patterns; the second is a test of similarities and differences, using pictures of familiar objects as materials; the third is a form board test, arranged so that it can be presented on paper; and the fourth is a test of absurdities, again using pictures.

[1] Pressey, L. W. "A group scale of intelligence for use in the first three grades," *Journal of Educational Psychology*, 10:297–308, September, 1919.

The tests have been given to about twenty-five hundred public school children (June, 1919) in the first three grades;[2] also to 64 "children" with a mental age on the Stanford-Binet Scale of between six and eight in the state school for the feeble-minded in Minnesota. Many of the school children tested had been previously "Bineted."

2. *Methods.*—The writer has used three methods for arriving at a conclusion concerning the validity of the scale. (1) One hundred and twenty children who had been "Bineted" were divided into five groups on the basis of their Intelligence Quotients (I. Q.'s) as determined by their scores in the Binet Scale (Stanford Revision). The scores of these groups in the Primer Test were then examined for consistency with the groupings according to the Binet ratings. (2) Three correlations with the Binet ratings were calculated: first, the correlation between the Binet "mental age" and the Primer Test score for the 64 institutional feeble-minded cases; second, a similar correlation for 148 children in the primary grades of Council Bluffs, Iowa; and third, a correlation for 57 unselected six-year-old children in the first grade at Washington, Indiana. (3) The writer chose this last group for a more detailed analysis by the method of partial correlation. The separate tests were intercorrelated, each test was correlated with the Binet ratings, and the partial coefficients and regression equation were calculated.

3. *Results.*—The results of the first method are indicated in Table I which gives the distribution of scores on the Primer Scale of 120 unselected six-year-old children, for whom the writer had Binet ratings. The cases have been divided into five groups on the basis of the Binet I. Q's. Group I (the first column) includes the scores of those children with an I. Q. of 125 or above; Group II contains those children who obtain an I. Q. of 110-124; Group III those with an I. Q. from 90-109; Group IV those from 76-89; and Group V those with an I. Q. below 76. The five groups might be termed, very superior, superior, average, inferior, very inferior.

There is some over lapping from one group to another, but the only distribution for which the amount of over lapping is surprising is that of the "average" group (Group III) whose distribu-

[2] The writer wishes to express her obligations to the school superintendents of Washington, Bedford, and Bloomington, Indiana, and of Council Bluffs, Iowa; also, to Dr. F. Kuhlmann, Psychologist of the School for the Feebleminded at Faribault, Minnesota for the materials here presented.

tion over laps that of every other group. The two extreme groups (I and V) are quite distinct from each other—that is, the lowest score in Group I is above the highest score in Group V. With the exception of one case, Groups I and II combined (i.e., the really bright children according to the Binet Scale) show no over lapping of test scores with the Groups IV and V combined (i.e., with the really dull children). The horizontal lines in Table I show the location of the 75-percentile, the median and the 25-percentile for the total distribution of these unselected children—i.e., the standard norms derived from the complete surveys made. As will be seen, the children in Group I make no scores below the 75-percentile for their age, while the children in Group V, with one exception, make no scores above the 25-percentile.

The lowest group is the most compact of the five. There is, in general, not as much over lapping of one group upon the other at the lower end of the distribution as at the upper end. This is understandable, as the brilliant children might not give their best performance and would thus make a lower score than they really should; while, on the other hand, it is not to be expected that adventitious circumstances would cause a dull child to be rated too high to the same degree. On the whole, the Primer Scale seems to have marked off degrees of ability as well as could be expected of a group test given to such young children. In this connection, it might be emphasized that some six-year-old children who make a low score on a group test receive a higher rating on the Binet examination because of the very different testing environment. This aids in explaining the dropping down of scores in the first three groups into the distributions of the lower groups. Some of the children making these scores lack, not intelligence, but independence. When asked to perform certain tasks in a group, they do not have the initiative to do their best; whereas they show a Binet rating more nearly commensurate with their true intelligence, because the examiner can constantly recall their attention and their efforts to the task in hand. The writer feels that the relationship shown in Table I is probably as high as can be expected between a group test and an individual test,[3] especially with such young children.

[3] Lowell, Frances. "A group intelligence scale for primary grades," *Journal of Applied Psychology*, 3:215-48, September, 1919. The scale described in this article is a group form of the younger years of the Binet; yet the correlation between group score and the rating on the Binet Scale when given to the same children individually is only +0.75.

TABLE I. THE DISTRIBUTION OF SCORES MADE BY 120 SIX-YEAR-OLD CHILDREN ON THE PRIMER SCALE, GROUPED ACCORDING TO I. Q. ON THE STANFORD-BINET SCALE

	Primer Scale Scores	Group I (I. Q. above 125)	Group II (I. Q. 110-124)	Group III (I. Q. 90-109)	Group IV (I. Q. 76-89)	Group V (I. Q. below 76)
		1	*2*	*3*	*4*	*5*
	80	1				
	78					
	76					
	74					
	72	2		1		
	70	1				
	68					
	66	1	2			
	64	1	1	1		
	62	4	4	2		
	60	3	2	1		
	58	2	2	3		
	56	4	1	4		
	54		1	4		
	52			5		
	50		1	1		
	48	2	2	1		
75-percentile						
	46		3	3	1	
	44			5		
	42		2	4		
	40		1	2		
	38		1	3		
Median						
	36		2	1		
	34			1		
	32			1		
	30			2		
	28			4		1
	26					
	24			2		
	22					
	20			1		
	18					
25-percentile						
	16			1		1
	14			2	2	
	12					
	10			1	2	
	8			1	2	
	6			1	1	1
	4					1
	2					1
	0					3
	No. cases..	21	25	58	8	8
	Medians	61	55	45	11	4

The second method of studying the validity of the Primer Scale involved the calculation of the Pearson coefficient of correlation between scores derived from it and scores obtained by using the Binet Scale. The crude coefficients of correlation for each of the three groups of children are as follows:

64 children from a feeble-minded school	+0.75
148 children in primary grades	+0.66
57 unselected children 6 years of age	+0.62

In all cases, the correlation is between group test scores and "mental ages" on the Binet examination.[4] The correlations are, the writer feels, as high as can be expected from the very different nature of the testing conditions for the two types of tests (group and individual).

The scores of the 57 unselected six-year-old children were subjected to further analysis by means of partial correlation. The crude coefficients of the scores of each of the four tests which make up the Primer Scale with each other and with the Binet ratings are given in Table II.

TABLE II. CRUDE COEFFICIENTS OF CORRELATION BETWEEN TESTS OF THE PRIMER SCALE AND THE BINET SCALE, TAKEN PAIR BY PAIR

Tests	1	2	3	4
1				
2	0.59			
3	0.48	0.46		
4	0.57	0.55	0.33	
5	0.53	0.45	0.42	0.47

LEGEND

1 = Test 1, Primer Scale
2 = Test 2, " "
3 = Test 3, " "
4 = Test 4, " "
5 = Binet Scale

This table should be read: the correlation between Tests 1 and 2 of the Primer Scale was found to be 0.59; between Tests 1 and 3, 0.48; etc.

It will be noticed that the coefficients of correlation between the tests of the Primer Scale are rather high, except between the third and fourth. Because of this difference in the relationship

[4] The I. Q., which is frequently used for such correlations, is a percent statement. No similar statement was readily possible with the Primer Scale. Therefore, it was necessary to use *total* scores on both scales. Even in correlating the scores of children of the same age, the error introduced in using I. Q. is considerable, as the basis for calculating the I. Q. may range (in the case above) from 6 years and 0 months to 6 years and 11 months—or nearly a whole year.

between the tests, it is especially desirable to find the partial correlations of each test with the Binet Scale, to determine just what the independent value of each test might be. The resulting partial coefficients of correlation are as follows:

$$r_{51.234}=0.2436$$
$$r_{52.134}=0.0951$$
$$r_{53.124}=0.1834$$
$$r_{54.123}=0.2006$$

These partial coefficients should be read: the correlation of 5 (Binet rating) with 1 (Test 1), the other three tests being constant, is 0.2436; of Binet rating and 2 (Test 2), the other three tests being constant, is 0.0951.

The partial coefficients are interesting in themselves aside from the part they play in calculating the regression equation.[5] The relative standing of the four tests in their relation to the Binet Scale, has not changed materially from that indicated by the original correlations (Table II). The first test still has the highest correlation, and the order of decreasing coefficients is now the fourth, the third, and the second according to the partial correlations shown above. The second test seems to contribute almost nothing to the correlation of the whole scale with the Binet Scale, which is not already contributed by some other test. When the other tests are held constant, its correlation with the Binet Scale becomes practically zero (0.0951). Thus, the second test is, according to this criterion, the least valuable test of the scale[6] and the first test is the most valuable.

III. The Reliability of the Primer Scale

1. *Materials.*—The papers of all the children in the first and second grades, and of all the six- and seven-year-old children in

[5] The regression equation of the Binet ratings on the four tests of the scale was found to be the following:

$$x_5=0.5719x_1+0.2142x_2+0.2844x_3+0.5380x_4,$$

where x is the deviation of a pupil's score in Test 1 from the average, x_2 his deviation from the average in Test 2, etc. x_5 is the most probable deviation from the average Binet score. The scores of the original 57 papers used in the derivation of the equation were weighted in accordance with the above equation. This weighting raised the correlation from 0.62 to 0.65.

[6] Since writing the above, the writer has tried a new method of scoring the second test, which takes account of possible chance successes, and has obtained a correlation of 0.52 instead of 0.45 as given in Table II. This difference in correlation might make some difference in the relative standing of the second test as revealed by partial correlation.

the city most recently tested were chosen as the material from which to estimate the reliability of the scale.

2. *Methods.*—The degree of the reliability of the measures obtained by the scale and of its separate tests has been calculated in two ways. Coefficients of reliability have been found for each of the four groups just mentioned (grades I and II and ages six and seven). As there was only one form of the scale, the coefficients of reliability were calculated by dividing the scale into halves (using alternate items), correlating the scores of the half tests, and applying Brown's formula.[7] This formula represents the extent to which the amalgamated results of two tests would correlate with a similar amalgamated series of two other applications of the same test. The formula is:

$$r_2 = \frac{2r_1}{1+r_1}$$

Both the total scale and each test have been treated in this way for the four age and grade groups. (2) The reliability has been calculated in terms of the probable error (P. E.), by the following method. Each pupil's score on one half of the scale was subtracted from his score on the other half. These differences were then averaged, giving an "average difference."

$$\text{The P. E. of a scale} = \frac{0.8453}{\sqrt{2}} \times \text{Av. Diff.}$$

$$= 0.5978 \times \text{Av. Diff.}^{8}$$

This formula applies whether the two tests from which the average difference is found are two separate equivalent tests or two halves of the same test. If two separate forms of the same test had been used, the above formula (P. E. = 0.5978 × Av. Diff.) would give the correct value, but in the case of the two halves of the same test, a further calculation must be made. The formula gives the P. E. of the half tests, but what is desired to obtain finally, is the P. E. of the *whole* test—not the P. E. of one of its halves. The P. E. of a test increases as the square root of its length; that is, in this case, the whole test is *twice* as long as its halves, and its P. E. consequently increases by $\sqrt{2}$. The P. E. of the whole test would, then, be equal to that of the

[7] Brown, William. *Essentials of mental measurement.* New York: G. P. Putnam's Sons, 1911, p. 101.

[8] Thorndike, E. L. *An introduction to the theory of mental and social measurements.* New York: Teachers College, Columbia University, 1916, pp. 190-93.

half test, multiplied by the $\sqrt{2}$. The resulting formula for deriving the P. E. of a test from that of its halves becomes:

$$\text{P. E.} = 0.5978 \times \text{Av. Diff.} \times \sqrt{2}.$$[9]

3. *Results.*—The coefficients of reliability for grades I and II and for ages six and seven are given in Table III. For grade I and age six, the first column represents the reliability of the scale, including all zero scores; the second column gives the reliability

TABLE III. THE COEFFICIENTS OF RELIABILITY BETWEEN THE THE TWO HALVES OF THE PRIMER SCALE FOR GRADES I AND II AND FOR AGES SIX AND SEVEN, AFTER APPLYING THE FORMULA:

$$r_2 = \frac{2r_1}{1+r_1}$$

(In this formula r_1 is the coefficient of correlation obtained by correlating the two halves of the scale and r_2 is the coefficient of reliability given in the table below.)

	GRADE I (142 CASES)		AGE SIX (90 CASES)			
	Zero Scores Included	Zero Scores Omitted	Zero Scores Included	Zero Scores Omitted	Grade II (120 Cases)	Age Seven (103 Cases)
	1	2	3	4	5	6
Total score	0.96	0.95	0.95	0.94	0.89	0.92
Test I	0.94	0.90	0.92	0.89	0.89	0.84
Test II	0.92	0.81	0.88	0.81	0.67	0.68
Test III	0.89	0.81	0.88	0.81	0.93	0.93
Test IV	0.82	0.78	0.84	0.79	0.54	0.75

omitting the zero scores. Neither method is satisfactory. If the test were to be given again, some but probably not all, of those children making zero scores would make a higher score. Since some of them would, the first column of coefficients, including all the zero scores, is too high: but, since not all of them would, the second column of coefficients, omitting all zero scores, is too low. Thus the true coefficient probably lies somewhere between the two.

The coefficients of reliability are uniformly high for the whole scale and for the first and third tests, but vary considerably for

[9] The writer is indebted to Dr. Walter S. Monroe for this method of determining the P. E. of a scale from the P. E. of its halves; as well as for many other helpful suggestions and criticisms in the course of the statistical analysis.

the second and fourth tests in the different groups. The reliability, as measured by the correlation of one half of a scale to the other, is in part a measure of the consistency of the performance of the pupils, and in part a measure of the consistency of the test from item to item. It is the judgment of the writer that the reliability coefficient is mainly an index of the latter factor. If the halves of a test correlate highly, it is evident that the items composing the two halves demand the same, or approximately the same kind of mental effort. For Tests I and III this consistency is high; these tests are both "form" tests, they demand no information of any kind, and it apparently makes little difference which set of items is used. Tests II and IV call for information of a practical sort; the two halves of these tests correlate rather poorly—sometimes very poorly. Evidently, practical information is not at all the unitary thing that form discrimination is; and the kind of information demanded by one item is different from that demanded by another. If still another form of the test were to be used, it is probable that these two tests would still correlate poorly, since these new items would involve still different informational elements.

The P. E. of the scale was derived in the manner explained above. The P. E. of grade 1 (142 cases) was 2.39 "points," or units; that of age six (90 cases) was 2.20 "points," or units on the scale. Since the difference between the median of one age and the median of the next age is, on the average, twelve "points," or units, a P. E. of 2.39 points would equal 2.39 months, and a P. E. of 2.20 points would equal 2.20 months. Again, since there are about twelve points from one age to the next, a statement, for example, that a given child scores at the median for the age above his own, does not seem to be invalidated by the unreliability of the scale, since the differences between the successive medians are well beyond the P. E. of the scores.

The P. E. of the Stanford-Binet Scale has been estimated by Otis to be about three and a half months.[10] The P. E. of the Primer Scale is even less. If the same reliability can be obtained with a group scale as with an individual scale, the group scale can,

[10] Otis, A. S. "An absolute point scale for the group measurement of intelligence," *Journal of Educational Psychology*, 9:341, June, 1918.

to some extent, take the place of the individual scale, at least for the preliminary sorting of pupils.

On the whole, the scale seems very reliable—unusually so when the ages of the children are considered. Considerable confidence may, therefore, be placed in the accuracy of the scores obtained.

IV. Summary

1. The writer presents data to show the reliability and the validity of the Primer Scale.

2. The validity is shown (1) by the distribution of scores made on the Primer Scale by children who have been classified according to I. Q. on the Binet examination; (2) by the correlations for various groups of children between the group scale and the Binet ratings; and (3) by partial coefficients of correlation.

3. The reliability is shown by (1) the coefficients of reliability for different ages and grades and (2) by the P. E. of the whole scale, as derived from the P. E. of the half scale.

4. It is concluded that the Primer Scale is reasonably valid and reliable.

THE APPLICATION TO TABLES OF DISTRIBUTION OF A SHORTER METHOD FOR COMPUTING COEFFICIENTS OF CORRELATION

LEONARD P. AYRES
Russell Sage Foundation

In the March number of the JOURNAL OF EDUCATIONAL RESEARCH there appeared an article by the writer on a shorter method for computing the coefficient of correlation. The method described avoids the necessity for finding the deviations of the separate measures from their averages and so eliminates the necessity for watching for unlike signs and making the series of corrections to allow for differences between the guessed averages and the true averages of the two series. The present article explains the application of the method to tables of distribution.

The new method is based on the principle that in any series of numbers the sum of the squares of the deviations from the average is equal to the sum of the squares of the numbers in the series, minus the product of the total of the series and its average. The formula for this method is as follows:

$$r = \frac{\text{Sum of products of subject and relative items minus av. of subject items} \times \text{tot. of relative items}}{\sqrt{\text{Sum of squares of sub. items minus av.} \times \text{total of these items} \times \text{Sum of squares of rel. items minus av.} \times \text{total of these items}}}$$

In using this method the multiplications are made first and the corrections are made separately. For two simple series of paired measures the operations may be illustrated as follows:

	Sub.	*Rel.*	*Sub.*^{2}	*Rel.*^{2}	*Sub.* $\times$ *Rel.*
	2	3	4	9	6
	3	2	9	4	6
	4	4	16	16	16
	7	4	49	16	28
	9	7	81	49	63
Tot.	25	20	159	94	119
Av.	5	4			

$$5\times20=100$$
$$5\times25=125$$
$$4\times20=\ 80$$
$$119-100=19$$
$$159-125=34$$
$$94-\ 80=14$$

$$r=\frac{19}{\sqrt{34\times14}}=\frac{19}{\sqrt{476}}=\frac{19}{21.82}=0.87$$

The procedure is similar when the method is applied to a distribution table. To illustrate it the case may be taken of a suppositious age and grade table for the elementary grades of a school system in which one hundred children enter the first grade each year at the age of six, are promoted by annual promotions of 80 percent so that eight in each ten go forward and two repeat, and none drop out of school during the elementary years. The problem is to compute the coefficient of correlation between the ages and the grades of the children. The work is as shown in the accompanying computations. The letters "S" and "R" stand for "subject" and "relative."

S	R–Ages										
Grades	6	7	8	9	10	11	12	13	T	ST	SST
1	100	20	4	1					125	125	125
2		80	32	9	2				123	246	492
3			64	39	15	5	1		124	372	1,116
4				51	42	20	8	2	123	492	1,968
5					41	42	24	11	118	590	2,950
6						33	41	27	101	606	3,636
7							26	39	65	455	3,185
8								21	21	168	1,344
T	100	100	100	100	100	100	100	100	800	3,054	14,816

RT	600	700	800	900	1,000	1,100	1,200	1,300	7,600	Av. of sub. items $\frac{3,054}{800}=3.818$ Av.
RRT	3,600	4,900	6,400	8,100	10,000	12,100	14,400	16,900	76,400	
ΣSf	100	180	260	340	422	503	583	666		Av. of rel. items
R(ΣSf)	600	1,260	2,080	3,060	4,220	5,533	6,996	8,658	32,407	$\frac{7,600}{800}=9.5$ Av.

$$3{,}054 \times 9.5 = 29{,}013$$
$$3{,}054 \times 3.818 = 11{,}660$$
$$7{,}600 \times 9.5 = 72{,}200$$

$$32{,}407 - 29{,}013 = 3{,}394$$
$$14{,}816 - 11{,}660 = 3{,}156$$
$$76{,}400 - 72{,}200 = 4{,}200$$

$$r = \frac{3{,}394}{\sqrt{3{,}156 \times 4{,}200}} = \frac{3{,}394}{\sqrt{13{,}255{,}200}} = \frac{3{,}394}{3{,}641} = 0.932$$

In the illustrative example the columns headed "ST" and "SST" give the items of the subject series and the squares of these items. They are found by the simple process of multiplying the numbers in the total column by those in the column headed "S" at the extreme left of the table and then multiplying these new totals by the "S" numbers once more. A similar process gives the "RT" and "RRT" rows at the bottom of the table.

The next step is to find the products of the paired values. This is done in the two last rows at the bottom of the table. The first is headed "ΣSf" and gives for each column the sum of the products of the numbers in the "S" column and their corresponding frequencies in the body of the table. Thus the sum of the products for the second column is 180 and consists of 1×20 plus 2×80.

After all these sums of products are found and entered in the row headed "ΣSf," they are multiplied by the "R" numbers at the heads of the columns and the results entered in the lowest row headed R(ΣSf).

The averages are found by dividing the total of the subject items (3,054) by the number of cases (800) which gives 3.818. Similarly the sum of the relative items (7,600) is divided by 800 to give the other average of 9.5.

The next step is to find the three "total × average" products and to make the three corrections which appear after the table. The first product is that of the total of one series (3,054) and the average of the other (9.5), or 29,013. This is subtracted from the sum of the products of the paired values (32,407) giving the remainder 3,394. This equals the sum of the x times y products as found in the conventional formula:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\ \sqrt{\Sigma y^2}}$$

Similarly the sum of the squares of the items of the subject series (14,816) is corrected by subtracting from it the total of the series times its average. The corresponding operation is done for the relative series. The final operations follow the formula given at the beginning of the article and are the same as those of the conventional procedure.

The work is greatly facilitated by the use of an adding machine. The writer did all the work of the example given in twenty minutes after the age-grade table had been written down and checked it by repeating all the operations in fifteen minutes. This was done with the aid of a comptometer—a small, non-listing, adding machine which can also be used for multiplying.

On work such as that given in this article the new method is found to be in practice much faster than the old one and errors in computation are very much less likely to be made.

Erratum

In Dr. Ayres' first article in the March number of the JOURNAL OF EDUCATIONAL RESEARCH at page 220, middle of page, read:

"Second correction $= 214 - (6 \times 30) = 214 - 180 = 34$."

Editorials

THE ACHIEVEMENTS OF THE SCHOOL YEAR

As the school year draws to a close, do the plans of superintendents and teachers involve taking an inventory of the achievements of the year? The merchant, the manufacturer, or the banker (and we are beginning to include the farmer) who does not at least once a year take an inventory of his business is generally considered to be headed toward certain failure. If when dealing with such tangible things as clothing, drygoods, machinery, automobiles, corn, cattle, and hogs, it is necessary to measure accurately the results of one's activities during the year, how much more is it necessary to render an accounting when one is dealing with the less tangible achievements of the school room.

In the records of the school census, attendance, enrollment, withdrawals, and promotion the superintendent has at hand much of the crude data which an inventory of his business requires. Many superintendents have given, at times during the year, one or more standardized tests. The results from these furnish additional crude data. However, in order to secure a satisfactory inventory, there should be an appraisal of the achievements of the pupils during the school year. The way to do this is to plan a comprehensive testing program for use toward the close of the school year. With the large array of standardized tests now available, the superintendent can secure the instruments needed for measuring many of the important achievements accomplished during the year.

The labor involved in the preparation of an inventory of the year's work will perhaps appall many schoolmen. Some will doubtless claim that they have been successful in the past without rendering an account of the year's work. In reply to these it is only necessary to say that there was a time when many merchants were successful who never made an inventory of their business.

But the time has now passed in the business world when a man can expect to be successful without it. If the time has not arrived in educational activities when such a procedure is necessary for success, the time is rapidly approaching. It, therefore, behooves those superintendents who desire to be rated among the successful ones to plan definitely for the rendering of an accounting of the year's work.

W. S. M.

THE GIFTED CHILD

For years serious attention has been given to the handling of backward and defective children. This involves methods of instruction, curriculum making, promotion, and segregation. Such efforts are worth while, and we do not wish to minimize their importance. We do wish, however, to insist that we must give more attention than has heretofore been given to the education of children of superior mental ability.

Superior children who are merely carried along through the grades under the customary "time service" spend many days of their school life during which their minds are being injured rather than educated. Many a boy, instead of letting his brain power run idle in the classroom while slower minds labor to understand problems that present no difficulty to him, would be better off playing in the field. Ought a bright child be penalized because he is bright? We speak of equal rights; but equal rights involve equal opportunities. And such children are robbed of their opportunity for self-realization.

Conservation of power and reduction of waste is an important social ideal. A lot of the brain power in our schools is daily going to waste because it is idling far below its normal working capacity. Genius which spends itself on inferior work is wasted. Our educational practice needs a severe earthquake that will shake us loose to the recognition of this waste.

Recently we have heard more about "doing something for superior children," but actual efforts to discover such children have been comparatively few and spasmodic. With the present development of mental testing we have an effective tool by which teachers and school administrators may make systematic search for these children. If such an instrument is not used, each gifted

child must force himself upon the attention of someone who has backbone enough to break away from tradition. Classroom procedure is not organized so as to favor a child's forcing himself into recognition. Moreover, in many places even if superior ability is perfectly evident, nothing short of a miracle will cause the traditional system to break its regularity of habit.

The latest development in mental testing is the group test. It can be given to hundreds of children in a single day. It affords impersonal data from which to estimate a child's capacity, and it thus removes the decision from the field where the personal feelings of teachers or of fond parents play a part. Experience proves that these group tests, in spite of the ease with which they may be administered, have a high degree of reliability. The results can be used for "spotting" individuals; and wherever there are serious differences between test results and other data, the child may be given further study or additional tests. These group tests ought to come into immediate and extensive use for the discovery of superior mental capacity in all classes from grade three up. Indeed, some group tests promise to be fairly reliable even below the third grade. Individual mental tests should be used on all school children who are too young to be tested by means of a group test.

No superintendent or principal should be satisfied with anything less than a complete and systematic search of his school for superior ability. Having found such ability, he should devise ways and means to handle the administrative problems of education so that this ability may be put to work. This is not altogether easy. It means that more thinking and more effort must be put into both administration and teaching. The problem can be solved, and it is being solved in many places. Various plans are being used but none of them is as important as the clear realization that superior mental capacity should be used and trained instead of being allowed to waste away in the formation of habits of idleness and mischief.

With such a realization, intelligence properly applied will meet the problem. Whether we like the work or not, it must be done. Let us face the task squarely. With due apology to advertising, we may say "do it now, for eventually you must."

VIRGIL E. DICKSON

Reviews and Abstracts

COURTIS, STUART A. *The Gary public schools: measurement of classroom products.* New York: General Education Board, 1919. 532 pp.

This is the last volume of the report of the survey of the Gary public schools and is based upon results of a number of educational tests given to the pupils in the Froebel, Emerson, Jefferson, and Beveridge schools between April 11 and June 9, 1916. The testing program was introduced by a series of nine simple psychological tests the scores from which are not included in the report. The purpose of giving these introductory tests was to familiarize the pupils with the procedure of testing. Each test was given throughout the four buildings on the same day except for occasional classes which were tested on the following day. Unusual care was taken to control the conditions of testing. Mr. Courtis and three trained assistants gave practically all of the tests. Automatic timers were used when feasible. The test papers were scored in duplicate. These precautions give the resulting scores a higher degree of reliability than they would have under ordinary conditions.

The tests and the grades to which they were given were as follows:

I. Spelling.
   - Cleveland List Tests, grades two to twelve
   - Courtis Dictation Spelling Tests, grades two to twelve
   - Spelling was also tested by means of the Denver Composition Test, grades four to twelve

II. Handwriting. The quality of the following types of material was measured by the Ayres' Handwriting Scale, "three slant edition."
   - Free Choice Handwriting Test, grades three to twelve
   - Denver Composition Test, grades four to twelve

III. Arithmetic
   - Courtis Standard Research Tests, Series B, Forms 3 and 4, grades four to twelve
   - Cleveland Survey Arithmetic Tests, Sets C, G, H, L, and O, grades three to twelve

IV. Reading
   - Kansas Silent Reading Test I, grades three to twelve
   - Kansas Silent Reading Test II, grades four to eight
   - Kansas Silent Reading Test III, grades eight to twelve
   - Gray's Oral Reading Test, grades three to eight
   - Reproduction Tests, grades two to twelve

Trabue Language Scales, B, C, D, and E, grades four to twelve

V. Language

Denver Composition Test, grades four to twelve

This volume stands out as distinctive in the series of volumes which make up the report of the survey of the Gary public schools. In the first place, it contains 532 pages besides the introduction and author's preface. The next largest, which is the summary volume, has only 265 pages. Second, it comes from the press several months after the others have been in the hands of readers and more than three years after the giving of the tests in Gary was completed. Third, the volume is a combination of the report of the survey and a treatise on the theory of educational measurements.

Besides a chapter given to each of the school subjects for which tests were given there are four other chapters: Chapter 1, "General Statement," in which is included a brief discussion of the status and limitations of educational measurements; Chapter 2, "Tests and Testing Conditions;" Chapter 8, "Factors Affecting Performance;" Chapter 9, "Conclusions." There is an effort to make the report as readable as possible. The presentation for each school subject is divided into three parts. First, there is a concise non-technical discussion of the significant aspects of the data secured. Second, there is a discussion of "all that is involved in the testing process." Third, in the appendices of the volume there is reproduced a number of the tabulations and a quantity of other data in which students of education will be interested. This organization materially facilitates the reading of the report for one who is not interested in critical discussions of the tests. One is disappointed, however, in not finding a concise summary statement of the findings for each subject. A partial exception to this statement should be made in the case of handwriting. A judicious use of paragraph headings would materially increase the clarity of the report.

The volume must be judged both as a report of the Gary survey and as a treatise on educational measurements. The conception of the survey is shown to have been comprehensive and the precautions exercised in its execution may well be taken as a model for future surveys. For example, several tests were used in each school subject. This is good procedure, because no single test can possibly have the same reliability that a battery of several tests has. Furthermore, no one test can cover the entire field of a school subject. Again, the scoring was done in duplicate by pupils and checked by examiners. This also contributes to the reliability of the published results. Among a number of other instances of the precautions taken, one may mention the fact that the judges who rated the compositions were carefully trained to their work before they began scoring.

In general the interpretation of the scores is made with unusual care and fairness. In English composition, for example, pupils wrote on the same subject and under the same conditions as obtained in the surveys of Denver and Grand Rapids. The compositions from these latter cities were, however, measured by the Willing Scale. In order to secure comparative data by which to interpret the Gary results it was obviously necessary to have the Gary papers rated on the same scale. The added precaution

was taken of having them (those of the eighth grade) rated by the same person—namely Mr. Willing—who rated the papers at Denver and at Grand Rapids. Many other instances might be given of the unusual care and fairness with which interpretations were made. Indeed, the author shows a disposition to set up and dispose of as many possible objections as he can. He seems to assume—with some truth we admit—that his conclusions will be assailed. His state of mind in this matter might be described as "nervous."

The author states as a general conclusion, "that the product of the class room teaching of the fundamentals is, at Gary, poor in quality and inadequate in amount; it approximates in character the product of the poorer conventional schools, and reveals in no particular the slightest indication that it has been affected either favorably or unfavorably by the enriched curriculum or other special features of the Gary schools. . . . . The entire investigation reveals many and consistent evidences of careless work, imperfectly developed habits and marked lack of achievement." (p. 384) A careful study of the results of the tests reveals many instances which confirm the above statement. In many cases the Gary scores are below comparative data from conventional schools. In some cases, however, the Gary scores compare favorably with the "available standards" and in a few cases comparative data are not used which would result in a more favorable though not necessarily reversed conclusion in respect to Gary.

For example, in interpreting the handwriting scores for quality comparative scores from Iowa, Cleveland, St. Louis, Grand Rapids, and Kansas City and the standards proposed by Freeman and Starch are used. It is entirely possible that the average performances of the pupils in conventional schools are too high for satisfactory standards. No reference is made to the studies by Ashbaugh[1] and Koos[2] in which they indicate that possibly the standards for quality should be lower than Freeman and some other investigators have indicated. Ashbaugh shows that the New York Civil Service Commission accepts 40 and 50 for "passing" in writing; and Koos proposes an ultimate standard of 60, even finding "considerable justification" for 50. The eighth-grade median for Gary was 43.

The conclusion is reached that "handwriting instruction at Gary is producing very small effect upon the product." (p. 46) This gives one the impression that the handwriting is entirely unsatisfactory. While it is true that quality at Gary is below even the lowest proposed standard, it is equally true that rate is notably above the rates used as comparative data. Rate of writing is important and credit should be given for this phase of the product of instruction in handwriting. Indeed, the author does so in other connections. It may be noted in passing that this is not the only instance in the report of inconsistency of statement.

[1] Ashbaugh, E. J. *Handwriting of Iowa school children.* (Bulletin of State University of Iowa, March, 1916.)

[2] Koos, L. V. "The determination of ultimate standards of quality in handwriting for the public schools," *Elementary School Journal*, 18:423, February, 1918.

Another example may be cited from the chapter on English composition. In addition to judging compositions by means of the Hillegas Scale, a study was made of the vocabulary used by the eighth-grade children at Gary. Comparison is made with the vocabularies reported by Jones and the conclusion is reached that "a careful study of the vocabulary fails to show any clear effect of special training at Gary." In the discussion there is no recognition of the possible effect of the foreign population of Gary upon the vocabulary of the children. This factor may have been taken into consideration but there is no direct evidence of it. Neither does he consider the possible effect of the fact that each child at Gary wrote but one composition—and that on "an exciting experience"—while the children in Jones' study wrote on the average 105 compositions each and were judged to have exhausted their vocabularies.

A third example of conclusions at variance with the sweeping general statement quoted above may be taken from the chapter on reading. The Kansas Silent Reading Tests yielded median scores at Gary which approximate the standard for conventional schools from the sixth grade to the eighth grade inclusive and which are rather notably above standard for grades nine, ten, and eleven. It is only in grades three and four that the Gary scores are materially below standard. It is true that these facts are stated by the author in their appropriate places. The point, however, is that they are not reflected in his general conclusions.

In the summary volume the following paragraph appears. "The impressions gained through inspection and the results of the tests are not entirely consistent. For example, silent reading makes a distinctly better showing in the tests than one would have expected on the basis of class room observation. Spelling makes a poor showing on the list test and a very good showing on the composition test. It is not possible to reconcile these divergencies without adducing considerations as to which different opinions could fairly be entertained. Without, however, seeking to ignore the conflict of evidence, the authors still feel that the quality of class room instruction at Gary falls short of what is necessary." (p. 102)

After studying the survey report on "Measurement of Classroom Products" and finding the interpretation of results open to such criticisms as indicated above, one cannot help feeling that at times the author has allowed his interpretations to be colored by what the writers of the summary volume suggest to have been the prevailing opinion of the survey staff. Even if such instances are not numerous in the report, it is unfortunate that they exist at all, because the objective nature of standardized tests has probably been the most convincing argument in favor of their use. If the measurement is objective but the interpretation becomes subjective, this argument loses its force.

As a treatise on the theory of educational measurements, the volume is a notable contribution. It contains a discussion of the nature of the abilities and detailed descriptions and analyses of certain of the tests used. Mr. Courtis gives for the first time an account of the construction of the Courtis Standard Research Tests in Arithmetic, Series B. A more complete

account is given of the validity and reliability of most of the tests used than has heretofore been published.

In Chapter VIII the author considers the factors upon which performance depends. This constitutes an important analysis, and throws an unexpected light on the meaning which may be attached to the Gary scores in different tests. The basic factor upon which performance depends is heredity or "capacity." The maturity of the pupil is a second important factor. Maturity together with general training (a third factor) is shown to be more potent in the Gary schools than specialized training (a fourth factor) in the development of the following abilities: rate of copying figures, oral reading, writing of multiplication combinations, handwriting (all tests), addition in Series B, and rate of silent reading as revealed by the Kansas Silent Reading Test. On the other hand, specialized training is more potent in engendering accuracy of performance—e.g., in silent reading, long division, addition and subtraction of simple fractions. This distinction is fundamental in judging the efficiency of specialized instruction on the basis of scores yielded by standardized tests—and it is a distinction which Mr. Courtis has more ably presented than any other survey writer.

Students interested in the theory of educational measurements will find the report a source of much valuable information. The various "critical discussions" of the tests, however, which are announced to contain "all that is involved in the testing process," fall disappointingly short of the realization of this ambitious announcement. For example, in the critical discussion of spelling, the author points out that the ability to spell is "specific" (although it may contain a general factor), and that from a child's ability to spell one word no reliable inference can be made concerning his ability to spell any other word. However, a pupil's spelling of a number of words can be used as a basis for infering his ability to spell other words belonging to the same difficulty group. The reliability of the inference is greater when random samples are used and increases with the number of words. Unfortunately, the discussion does not deal with the relation between the reliability of the inference and the number of words used, except to say, "it seems to the writer probable that the conclusions would not be changed were the number of test words to be largely increased."

Neither does the author investigate the relation between the reliability of a test and the difficulty of the words used, or the relation between the reliability of a test and the method of giving the test words. These omissions occur in spite of the author's statement that "the preceding discussions have had for their purpose the full statement in regard to the validity of the performance of an individual in a single test as a measure of his ability in spelling."

The discussions for the other subjects have similar limitations. At the present stage of the development of measuring instruments and the analysis of the measuring process, it is not to be expected that the "critical discussions" would be complete. In fact, they contain much more than one might expect, since they are somewhat incidental to a survey report. The propriety of attempting to include in such a report a discussion of "all that is

involved in the testing process" will doubtless be questioned by many readers. One finds many pages that appear to contribute very little to a report of the "measurement of class room products," although undoubtedly they will be commended by those who are interested in the theory of educational measurements and would very properly find place in a treatment of that topic. The inclusion of these topics undoubtedly delayed the preparation of the report, and in this way lessened the value of the volume as a report of the survey.

In the dual character of the report, there is evidence that as yet we have not found a satisfactory procedure for reporting the use of educational tests in a school survey. If the author gives no more than the results and his interpretation of them, he may be accused of misinterpretation and even misrepresentation by some of those who do not understand the process of measurement and the limitations of the particular tests used. On the other hand, if, as in this case, the author enters into a critical discussion of the tests and the theory of measurement, the report may lack unity and will certainly become voluminous. It may fail to be effective in making clear the actual conditions which were found to exist. Presumably Mr. Courtis was interested in the opportunities that the survey afforded for research. At any rate, it is certain that as a contribution to the literature of educational research his book is a more notable contribution than it is as a report of the measurement of classroom products at Gary.

W. S. M.

BUCKNER, CHESTER A. *Educational diagnosis of individual pupils* (Teachers College, Contributions to Education, No. 98). New York: Teachers' College, Columbia University, 1919. 93 pp.

FRETWELL, ELBERT K. *A study in educational prognosis* (Teachers College, Contributions to Education, No. 99). New York: Teachers' College, Columbia University, 1919. 55 pp.

These are companion volumes in the sense that the investigations which they report were carried out upon the basis of the same tests of the same boys.

Dr. Buckner investigates the methods by which scores in one test may be legitimately compared with scores in another, the amount and distribution of individual variability, the existence of erratic scores, the causes of extremely variant scores, and the relation between ability and variation. To express these scores in terms of the standard deviation or the quartile deviation permits comparison between them, and at the same time retains the refinement of the original scores.

As usual it was found that pupils varied widely in a given test, extreme scores being more numerous below the median than above it. It was also found that a given pupil differed markedly in his scores on different tests, the scores having first been made comparable. It appears that the more capable pupils—if we may draw practical conclusions from theoretical discussions—not only do better work but that they do it more consistently.

It is pointed out, moreover, that while some of the variability and some of the extreme scores are caused by distractions and other chance occurrences,

these factors are not as effective in producing high scores as in producing low scores. The practical meaning of this would seem to be that pupils who obtain extremely low scores are entitled to the benefit of further investigation.

It seems, therefore, to be clear that the determination of the status of a pupil must be carried out on the basis of a considerable amount of information concerning each pupil. This is the reason for teams or "batteries" of tests.

As to "Causes of Extreme Variability," the following are considered: (a) the nature of the tests; (b) their administration; (c) accidental occurences; (d) statistical treatment of the results; and (e) the ability of the pupil in different traits.

Two important practical considerations as to the nature of tests are distinguished. First, the range of a test must be somewhat greater than the range of ability in the group tested. Second, the intervals between possible scores on the test must be small.

We shall not take up each of these five causes. Naturally it is the ability of the pupil which most affects extreme variability. One of the practical conclusions reached in this connection is that extreme scores are little, if any, more unreliable than those near the median. This appears to be particularly applicable to low scores despite the fact that accidental occurences tend to depress rather than raise results. The evidence on this is the fact that ability and variability bear a direct relation to each other.

Dr. Fretwell takes up the investigation at the point where Dr. Buckner leaves it. He asks: "Can academic success be predicted? If so, how?" The conclusion is reached that academic success can be predicted. When all temporary illnesses, ranging from bad colds through contagious diseases to a month in the hospital, all fortunes or misfortunes in the home life, barring withdrawal from school, all the change in physical conditions and varying interests among boys of eleven to thirteen—when all these and a score of others that might be enumerated are considered—the use of these tests in one and one quarter hours' testing at the beginning of the year would have agreed with the classification of the teacher after teaching the pupils one year in 90 percent of all the cases.

As to the best basis of prediction, three others besides the tests are considered—school marks from the beginning of each pupil's school career, age, and the combined estimates of departmental teachers. The results of these comparisons lead the author to say: "In this study, academic success in the first year of junior high school was more successfully predicted by a group of standardized tests than by all previous school marks or age or teachers' estimates."

Some tests prove more valuable for prediction than others. The following is the order of importance as the author found it: reading (Thorndike's Understanding of Sentences), visual vocabulary, opposites, spelling, completion tests, arithmetic tests (Woody), easy directions, mixed relations, and composition.

B. R. B.

WILSON, GUY MITCHELL. *A survey of the social and business usage of arithmetic.* (Teachers College. Contributions to Education, No. 100). New York: Teachers College, Columbia University, 1919. 62pp.

The content of school subjects is largely determined by tradition. This simply means that the content as originally determined tends to be retained. Changes are made from time to time but in general they lag far behind social and school needs. Recently the attention of educators has been directed to curriculum making, and many investigations have been undertaken to determine what should constitute the curriculum in a number of school subjects.

Tradition has influenced to a very large degree the curriculum of arithmetic as it is now taught. Since the work of Rice in 1893 and McMurry's proposals enunciated in 1904, a number of investigators have attacked the problem of determining what the course of study in arithmetic should be. Professor Wilson has employed a procedure which is different from that used by preceding investigators.

He states that the "problem set for solution in this study is to determine the arithmetic actually used by adults in their social and business relations." Accordingly pupils in the sixth, seventh, and eighth grades in 23 cities, and in certain rural schools were requested to secure from their parents the arithmetical problems which the latter had solved during a period of two weeks. Reports were obtained from 4,068 persons and the total number of problems after unsatisfactory reports were rejected was 14,583. The author demonstrates that the group of persons reporting is probably sufficiently representative of the population of the entire country with the exception of communities which have specialized activities.

The 14,583 problems were analyzed as to the topics involved and the magnitude of the numbers occurring in the four fundamental operations. For both integers and fractions, common and decimal, the author found that six topics (namely, multiplication, addition, subtraction, division, fractions, and accounts) include slightly more than 90 percent of all the problems. The remaining 10 percent are divided among 37 additional topics.

With reference to the magnitude of the numbers, the author found that they rarely exceeded four places and that the majority of them did not extend beyond two places. He found only 50 different common fractions in all of the problems reported, and 40 of these occurred eight times or less.

Although the title and the explicit statement of the author's problem indicate that the study was made to determine what arithmetic is actually used by adults in their social and business relations, it appears from a reading of the monograph that the author's real purpose was to determine in a general way what should constitute the curriculum in arithmetic in the elementary school. In making this application of his results, he announces two main assumptions:

"(1) *Aim of Arithmetic in the Grades.* While not denying the cultural and disciplinary value of arithmetic—in common with any subject systematically studied and well taught—it is assumed that arithmetic in the grades is justified only on the basis of its utility in the common affairs of life. We learn the multiplication table not to sharpen the wits nor to comprehend a

beautiful system but to figure our bills, our taxes, or the interest on a note. Whatever arithmetic is given in the grades beyond the essentials required by social utility consumes time that could be used more profitably in other ways.

"(2) *Determining the Social Utility of Arithmetic.* Social utility can be determined only on the basis of the uses of arithmetic in adult activities."

Professor Wilson would eliminate decimal fractions except in United States money, proportion which he states is really absent, square root, reductions in denominate numbers, exchange, and the metric system in addition to eighteen other topics. He believes it to be safe and conservative to recommend the omission of these topics. He says, "If to the four fundamentals and fractions, one were to add accounts, simple denominate numbers, and percentage, little would be left for all the other processes. So little, in fact, that it seems unfair to give attention to them as drill processes in the elementary schools. Some of them should receive no attention.

It appears that the author defines the use of arithmetic as "figuring" and that three additional assumptions are implied: (1) that adults are now making the uses of arithmetic which they should make in their social and business activities; (2) that the needs for "figuring" will be approximately the same in the future as they are now for the adults included in this study; (3) that the best way to prepare a child for "figuring" is to use a curriculum based solely upon the problems which adults use.

The author does not consider the school needs which a pupil may have for arithmetic in the study of geography, history, science, manual training, and other school subjects. Until the curriculums for these subjects are changed radically, pupils will have need for certain aspects of arithmetic. In the absence of a scientific analysis of their arithmetical requirements, one should not assume that they are included in the adult social and business usages of arithmetic.

Doubtless many readers will be unable to agree with Professor Wilson in his assumption that no arithmetic should be taught in the elementary school beyond the "essentials required by social utility" and that "social utility can be determined only on the basis of the uses of arithmetic in adult activities." Doubtless most readers will be willing to accept this as one of the important bases; but absolutely to reject other bases—as the author does—is to take a radical position. This is especially true since he implicitly makes other assumptions, as has been pointed out, which deserve consideration in the making of a course of study.

To define the usable product of arithmetical instruction as "figuring" is to limit decidedly the significance of the subject. It is true that "figuring" is the most tangible and conspicuous outcome; but other outcomes, although they are more subtle, are nevertheless important and should be definitely planned for in the making of a curriculum. Furthermore, it is not safe to build a curriculum on the assumption that adults are now making all the use of arithmetic which they should make in their social and business activities. There seems to be considerable unanimity of opinion that in many respects adults are now lacking in efficiency with respect to quantitative matters. As an instance of this we may cite the agitation for family budgets and farm accounting.

Professor Wilson's study will no doubt have a far-reaching effect upon school practice. It should, because it is by far the most valuable work which has been done on minimum essentials in the field of arithmetic. As a survey of the social and business usage of arithmetic the monograph is a notable contribution and will probably rank, as the author suggests, with that of Ayres and Jones in the field of spelling. To the writer it appears, however, that he has lost sight of certain important considerations in applying the results of his survey to the making of a curriculum for arithmetic, as is indicated in the implied assumptions given above. Some of his conclusions with respect to the curriculum will probably stand when a more conservative and analytical interpretation is made of the data collected, but all of them should not be accepted until the effect of the implied assumptions and possibly of other factors is known.

W. S. M.

Bibliographies

Standardized Tests for the High School

(Continued from March number page 242)

C. Algebra *(Continued)*

Stromquist's Preliminary Algebra Tests. This series includes tests upon the following operations: (1) addition; (2) subtraction; (3) multiplication; (4) division; and (5) factoring.

Publisher: University of Wyoming, Laramie, Wyoming.

Thorndike's Algebra Test. This is a series of eight exercises arranged in order of increasing difficulty as determined by the opinion of competent judges.

Reference: Thorndike, E. L. "An experiment in grading problems in algebra," *Mathematics Teacher*, 6:123-34, March, 1914.

D. Geometry

Irwin's Tests in Mental Manipulations of Space Relations. The author feels that this ability is not confined to mathematics alone, but it is constantly in demand in other lines as sewing, general science, drawing, English, etc. He has, therefore, worked out four tests (Tests A-D) which should "yield a rough measure of the ability of each pupil in a class to image space relations."

Address: H. N. Irwin, Fairmount Junior High School, Cleveland, Ohio.

Reference: Irwin, H. N. "Preliminary attempt to devise a test of the ability of high school pupils in the mental manipulations of space relations," *School Review*, 26:600-6, 654-70, 759-72, October-December, 1918.

The derivation of the tests is described and the tests are reproduced.

Minnick's Geometry Tests. This series of tests is based on the assumption that demonstration of a geometrical theorem involves the following

abilities: (1) the ability to draw the figure; (2) the ability to state the hypothesis and conclusion; (3) the ability to recall facts concerning the figure; (4) the ability to select and organize facts so as to produce the proof.

Publisher: J. H. Minnick, University of Pennsylvania, Philadelphia, Pennsylvania. Also obtainable from Bureau of Educational Research, University of Illinois, Urbana, Illinois.

Reference: Minnick, J. H. "A scale for measuring pupils' ability to demonstrate geometrical theorems," *School Review*, 27:101-9, February, 1919.

Study based on tests given in thirty high schools throughout the country and ranging in size from a few hundred pupils to several thousand. These pupils had completed either the first two books of plane geometry or all of plane geometry. Illustrated by graphs and tables.

Minnick, J. H. "Certain abilities fundamental to the study of geometry," *Journal of Educational Psychology*, 9:83-90, February, 1918.

The author discusses the four abilities which are fundamental to the study of geometry; how are these abilities related to teaching; to what extent are these four abilities developed; development of a series of tests which may be used for the purpose of diagnosis.

Minnick, J. H. *An investigation of certain abilities fundamental to the study of geometry.* Philadelphia: University of Pennsylvania, 1918.

In this monograph Dr. Minnick gives an analysis of geometrical ability and a complete account of the derivation of his tests.

Starch's Geometry Test.

Address: Daniel Starch, University of Wisconsin, Madison, Wisconsin.

Stockard and Bell Geometry Test. A test of seventy questions which involve drawing figures, naming figures, indicating order of development in demonstration, completing statements, stating the converse, definitions, regular polygons, parts of a demonstration, etc.

Stockard, L. V. and Bell, J. Carleton. "A preliminary study of the measurement of abilities in geometry," *Journal of Educational Psychology*, 7:567-80, December, 1916.

Derivation, preliminary form reproduced. Analysis of results.

IX. Mental Tests

Army Tests. The mental tests which have been used extensively in the army were prepared by the Psychology Committee of the National Research Council. Three systems of tests are now in use:

1. *Alpha.* This is a group test designed for men who can read English. It requires only fifty minutes to give. The exercises are designed so they may be answered without writing, merely by underlining, crossing out, or checking. The papers may be scored by means of stencils, so that nothing is left to the personal judgment of those who do the scoring.

2. *Beta.* This is a group test for foreigners and illiterates. Success in Beta does not depend upon knowledge of the English language, as the instructions are given entirely by pantomime and demonstration.

3. *Individual Tests.* Three forms of individual tests are used: Yerkes-Bridges Point Scale, the Stanford-Binet Scale, and the Performance Scale.

Address: Division of Psychology, Medical Department, War Department, Washington, D. C.

References: Army mental tests, methods, typical results, and practical application. Washington, D. C., November, 1918.

This bulletin describes the three types of tests used in the army. It also gives an explanation of letter ratings and directions for the use of intelligence ratings.

Madsen, I. N. and Sylvester, R. H. "High school students' intelligence ratings according to the alpha army test," *School and Society*, 10:407-10, October 4, 1919.

Results of the Alpha test given to the high-school students of Rockford, Illinois, Madison, Wisconsin, and Sioux City, Iowa.

Van Wagenen, M. J. "Our schools as measured by the army tests," *Educational Administration and Supervision*, 5:163-76, April, 1919.

Mr. Van Wagenen states: "The use of the Army Tests in certain elementary schools, high schools, and universities for the purpose of gathering information for the use of psychologists in the army has, at the same time, revealed very clearly certain facts about the schools, the students tested, and possible values of mental tests in educational practice." The author describes Army Test E and compares it with Alpha, form 6. Table II gives the weighted scores of elementary school pupils, high school freshmen, and college men and women.

Materials from the Army Tests under the title of "National Intelligence Tests" have been standardized by a committee of the National Research Council. They are to be published this spring, 1920, by the World Book Company, Yonkers, New York.

Holley's Sentence Vocabulary Scale. This scale has been devised with the Terman Stanford-Binet Vocabulary as a basis. The words of this well-known vocabulary are put in sentences, the last word of which is found among four words to the right. The pupils indicate the correct responses by underlining the word which completes the sentence. This scale has been found to be a fair measure of intelligence. Series 3B is for grades VII to XII.

Publisher: Bureau of Educational Research, University of Illinois, Urbana, Illinois.

Reference: First annual report of Bureau of Educational Research, 1918-1919.

Otis' Group Intelligence Scale. This scale is composed of ten tests which measure various phases of mental ability. Two forms of the scale are available. They may be used in grades VI to XII.

Publisher: World Book Company, Yonkers, New York.

References: Otis, A. S. "An absolute point scale for the group measurement of intelligence," *Journal of Educational Psychology*, 9:239-61, 333-48, May-June, 1918.

Report of the Department of Research, being part of a report of the superintendent of schools of Oakland, California, 1917, p. 210.

Theisen and Fleming's Classification Test. A group test based on the Army Test (Alpha), has been successfully used for high-school purposes. Takes fifty minutes to give and may be scored at the rate of 15 papers per hour.

Publisher: Bureau of Publications, Teachers College, Columbia University, New York City.

Reference: An article by Dr. Theisen entitled, "Group Intelligence Tests," will appear in an early number of the *Journal of Educational Research*.

Thurstone's Psychological Examination for College Freshmen and High School Seniors.

Address: L. L. Thurstone, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

X. Music

Baldwin's Public School Music Test. A preliminary test of the efficiency of public school music. It is aimed to measure the accomplishment of pupils. In this respect it is similar to the Courtis Standard Research Tests in Arithmetic.

Reference: Baldwin, Ralph. "Efficiency in school music teaching and practical test of the same," *Music supervisors' national conference proceedings*, 1914, p. 43-50.

On page 46 the test is reproduced.

Seashore's Musical Talent Chart. This chart is based upon the analysis of musical ability and offers a graphic means of representing the pupil's musical ability. The pupil's musical abilities are measured by his response to tones which are mechanically produced. The chart is simply a convenient device for recording the responses of the pupil. The records for testing a child are made by the Columbia Graphophone Company, Woolworth Building, New York. The *Manual for Measures of Musical Talent*, furnished free by the Educational Department of the Columbia Graphophone Company, contains directions, norms, and interpretations.

References: Seashore, C. E. "The measurement of musical talent," *Musical Quarterly*, 1:129-48, January, 1915.

A discussion of the elements which make up musical talent and the derivation of the Musical Talent Chart, the chart itself, and its meaning and use.

Seashore, C. E. "The rôle of a consulting supervisor in music," *Eighteenth Yearbook of the National Society for the Study of Education*, 1919, part II, p. 111-23.

Account of the use of the test in the elementary schools and with adults. Mr. Seashore states that the best time to use the chart is in the fifth grade.

Seashore, C. E. *The psychology of musical talent*. Boston: Silver Burdett and Company, 1919.

The system of evaluation of musical talents is reviewed and a chapter is devoted to each of the five double disk records.

Seashore, C. E. *Vocational guidance in music*. (University of Iowa Monographs, first series, no. 2, September, 1916.)

Copies may be obtained on request from the librarian of the University of Iowa, Iowa City, Iowa. A discussion of the value of the Musical Talent Chart.

"Testing for musical talent," *Musician*, 24:10, July, 1919.

A discussion of Seashore's tests.

XI. Physical Training

Athletic Badge Tests for Boys. These tests are not a complete measure of physical efficiency, but they serve as a fair index of heart, lung, and muscular development. They comprise a sixty-yard dash, standing broad jump, and a pull up. Standards have been worked out. These tests have been used in the high schools of Gary, New Orleans, Seattle, Buffalo, and New York.

Publisher: Playground and Recreation Association of America, 1 Madison Avenue, New York City.

Reference: Hanmer, L. F. *Gary public schools: physical training and play.* New York: General Education Board, 1919.

Athletic Badge Tests for girls. This test is similar to the one for boys, but being a newer series has no standard.

Publisher: Playground and Recreation Association of America, 1 Madison Avenue, New York City.

Reference: Hanmer, L. F. *Gary public schools: physical training and play.* New York: General Education Board, 1919.

Mr. Hanmer states: "The results (for the girls) are therefore no more satisfactory than the results of the boys' test. There is also the same unevenness and irregularity in development."

Baldwin's Physical Development Scale. Separate scales are given for boys and girls. Each scale gives norms for height, weight, and breathing capacity for each year from the kindergarten to the last year of the high school. This scale is included in Rapeer's Scale for Measuring Physical Education.

Address: Bird T. Baldwin, University of Iowa, Iowa City, Iowa.

Reference: Baldwin, B. T. "A measuring scale for physical growth and physiological age," *Fifteenth Yearbook of the National Society for the Study of Education*, 1916, part I, pp. 20-21.

Rapeer's Scale for Measuring Physical Education. This is a score card for judging five aspects of the results of physical education, i.e., health, physiological efficiency, physical development, physical ability and mental qualities.

Address: L. W. Rapeer, 1719 H. Street N. W., Washington, D. C.

XII. Reading

Kansas Silent Reading Tests. Test III of this series is designed for the high school. It consists of a series of exercises in which the pupil is asked to read a short paragraph and answer a single question on it. The administration of the test is very simple and it has been very widely used.

Publisher: Bureau of Educational Measurements and Standards, Kansas State Normal School, Emporia, Kansas.

References: Kelly, F. J. *The Kansas Silent Reading Test* (Studies by the Bureau of Educational Measurements and Standards, No. 3), Emporia, Kansas: Kansas State Normal School, 1915.

This contains the complete account of the derivation of the test.

Kelly, F. J. "The Kansas silent reading test," *Journal of Educational Psychology*, 7:63-80, February, 1916.

This is a partial account of the derivation of the test.

Monroe, Walter S. "A report on the use of the Kansas silent reading tests with over one hundred thousand children," *Journal of Educational Psychology*, 9:600-8, December, 1917.

The author concludes that the Kansas tests place more emphasis upon comprehension than upon rate of reading.

Smith, B. M. "Correlation of ability in reading with the general grades in high school," *School Review*, 27:493-511, September, 1919.

In trying to determine the correlation between abilities in reading and school abilities in the Oak Park, Illinois, high school Miss Smith used the Kansas Silent Reading and three other tests which she devised. Correlations are given for Kansas Silent Reading as well as the other tests.

An educational survey of Janesville, Wisconsin. Madison, Wisconsin: State Department of Public Instruction, 1918, p. 261.

An account of the use of the Kansas Silent Reading tests in high school.

Monroe's Standardized Silent Reading Tests. This series of silent reading tests is essentially a revision of the Kansas Silent Reading Tests described above.

Publisher: Bureau of Educational Research, University of Illinois, Urbana, Illinois.

Reference: Monroe, Walter S. "Monroe's standardized reading tests," *Journal of Educational Psychology*, 9:303-12, June, 1918.

This is an account of the derivation of the revision. The Kansas Silent Reading Tests have been criticized because many of the tests resembled arithmetical puzzles rather than ordinary reading material. In the revision the paragraphs have been selected from school readers.

School report of the town of Southington, Connecticut, 1919, p. 17.

An account of the use of these tests in the high school.

Neher's High School Vocabulary Scale. A test consisting of one hundred words chosen from the top of every third column of Laird and Lee's *Vest Pocket Dictionary*.

Address: H. L. Neher, Northwestern University, Evanston, Illinois.

Reference: Neher, H. L. "Measuring the vocabulary of high school pupils," *School and Society*, 8:355-59, September 21, 1918.

Reproduction of the test, scores obtained through experimentation and conclusions reached.

St. Paul Survey Silent Reading Scale for High School. This scale is similar to Minnesota Scale Beta, the derivation of which is based on Thorndike's Scale Alpha 2.

References: Report of a survey of the school system, St. Paul, Minnesota, 1917, p. 375-401.

The tests and questions are reproduced. Methods of scoring and a comparison of scores is also given.

Haggerty, M. E. and Thomas, M. J. "Preliminary study of the reading attainments of college freshmen," *School and Society*, 6:230-38, August 25, 1917.

The test was given to the college freshmen and was also given to two hundred and fifty members of the senior class in the St. Paul high school. The test which is reproduced in this article includes a number of the exercises from the St. Paul Survey Test.

Smith's **Silent Reading Paragraphs for High Schools.** Miss Smith has devised three informal silent reading tests. The tests are as follows:

Test I. Paragraph Test. Ability to grasp central idea. Four paragraphs were chosen from "A Father to His Freshman Son in College," (*Atlantic Monthly*).

Test II. Information Test. Ability to recall items of information. Selections were taken from the *Independent*.

Test III. Narrative Test. Ability to follow plot. Selection was a story from an inside page of *Chicago Daily Tribune*.

Address: Bertha M. Smith, Oak Park High School, Oak Park, Illinois.

Reference: Smith, B. M. "Correlation of ability in reading with general grades in high school," *School Review*, 27:493-511, September, 1919.

The tests were given to determine the correlation between abilities in reading and school abilities. The four reading tests used are described and the author's three tests are reproduced.

Starch's **English Vocabulary Test.** This test is designed to measure the percent of words of the entire English vocabulary as well as the absolute number of words that a person understands. The test consists of several sets of one hundred words which were selected at uniform intervals from the entire English vocabulary—Webster's *New International Dictionary*.

Address: Daniel Starch, University of Wisconsin, Madison, Wisconsin.

Reference: Starch, Daniel. *Educational measurements*. New York: Macmillan, 1916, p. 38 ff.

The test is reproduced. Tentative scores are shown for different grades in the high school as well as for elementary school and college.

Thorndike's **Scale Alpha 2 for Measuring the Understanding of Sentences.** This scale consists of a series of paragraphs and questions about the paragraphs arranged in order of difficulty of comprehension. This scale measures only the degree of comprehension and does this in terms of difficulty of the exercise which the pupil is able to do 80 percent correctly. Part II may be used in grades VI to XII.

Publisher: Bureau of Publications, Teachers' College, Columbia University, New York City.

References: Thorndike, E. L. "Improved scale for measuring ability in reading," *Teachers College Record*, 16:445-67, 17:40-67, November, 1915, January, 1916.

The test, key and score sheet are reproduced in the November issue along with the discussion of the use of the scale method of scoring, and the interpretation of scores. The derivation of the test is in the November number.

Kelley, T. L. "Thorndike's reading scale alpha 2 adapted to individual testing," *Teachers College Record*, 18:253-60, May, 1917.

Directions for using the scale with individual rather than with the class as the author intended.

XIII. Science

A. General

Caldwell's **Science Tests.** The tests may be divided into three groups. Group I consisting of two pairs of tests deals with observation and discrimination. Group II relates primarily to the pupil's ability to recog-

nize natural phenomena, to recall experiences in science study and to make constructive statements regarding these experiences. Group III consists of two tests of ability to give reasons for, to interpret, or to explain changing phenomena—the changes occurring in the presence of pupils in the form of an experiment or demonstration. These tests were used in the senior classes of the Gary, Indiana high school.

Reference: Caldwell, O. W. *The Gary public schools: science teaching.* New York: General Education Board, 1919.

Chapter X is devoted especially to the reproduction of the tests which have been prepared by Mr. Caldwell.

Downing's Information Tests in Science. A vocabulary test of one hundred scientific terms covering the entire field of high-school science. The pupils are asked to put the letter "E" before a word he can explain, and the letter "F" before a word which he has heard about.

Address: Elliot R. Downing, University of Chicago, Chicago, Illinois.

Reference: Downing, E. R. "A range of information tests in science," *School Science and Mathematics*, 19:228-33, March, 1919.

The author reproduces the list of one hundred terms and states results with about fifteen hundred students in seventh and eighth grades, high school, and university.

Grier's Range of Information Test in Biology.

Reference: Grier, N. M. "Range of information test in biology," *Journal of Educational Psychology*, 9:210-16, 388-93, April-September, 1918.

The design of the tests is that described by Whipple in his Range of Information Test. The first number contains a description of a test in physiology and the second one in zoology.

Herring's Tests in Scientific Thinking. Series of three tests on each of eleven abilities involved in scientific thinking.

Reference: Herring, J. P. "Measurements of some abilities in scientific thinking," *Journal of Educational Psychology*, 9:535-58, December, 1919.

The tests are reproduced and results of their application to one pupil is given.

Ruch's Range of Information Test in General Science. This test consists of fifty scientific terms assembled upon the basis of frequency in general science texts. As three series of words are given, a different series may be given during the year.

Address: G. M. Ruch, School of Education, University of Oregon, Eugene, Oregon.

Reference: Ruch, G. M. "A range of information test in general science," *General Science Quarterly*, 4:257-62, November, 1919.

Describes the derivation of the test which is reproduced. The method for scoring is discussed.

B. Chemistry

Bell's First-Year Chemistry Test. A test of 25 questions which the author considered basic for elementary chemistry.

References: Bell, J. C. "Study of the attainments of high school pupils in first-year chemistry," *School Science and Mathematics*, 18:425-32, May, 1918.

Brief account of its derivation. Results from 401 pupils analyzed.

Bell, J. Carleton. "A test in first year chemistry," *Journal of Educational Psychology*, 9:199-209, April, 1918.
This is the complete report of which the above is a preliminary statement.

Jones' Chemistry Tests. (Union Science Series, *See* Physics.)
Address: F. T. Jones, 10109 Wilbur Avenue, Cleveland, Ohio.

Rivett's Time Limit Test in Chemistry. The article presents two types of simple tests which are arranged for beginners. They are intended as suggestive types of tests which teachers of chemistry may easily devise. The pupils may score these tests themselves.
Address: B. J. Rivett, Northwestern High School, Detroit, Michigan.
Reference: Rivett, B. J. "Testing results in chemistry," *School Science and Mathematics*, 19:742-45, November, 1919.
Two types of tests are reproduced. Advantages over the old form examination are stated in conclusion.

Webb's Preliminary Test in Chemistry. Test requires pupils to classify fifty well-known substances as elements, compounds, or mixtures.
Address: Hanor Webb, George Peabody College for Teachers, Nashville, Tennessee.
Reference: Webb, Hanor. "A preliminary test in chemistry," *Journal of Educational Psychology*, 10:36-43, January, 1919.
This is an account of the use of the author's test with a number of his classes.

C. Geography

Boston Geography Test. A forty-minute test on United States and the countries of Europe. Although the tests were given to elementary, high, and normal school pupils, particular interest centers in the results achieved by the eighth grade.
Address: Leonard O. Packard, Boston Normal School, Boston, Massachusetts.
Reference: Geography. A report on a preliminary attempt to measure some educational results. (Bulletin No. 5 of the Department of Educational Investigation and Measurement, Boston Public Schools, 1916.)
Discussion of the tests with reproduction, but method of giving and of rating is not shown. Results for high-school students are given.

D. Physics

Chapman's Physics Test in Electricity and Magnetism, Sound and Light.
Address: J. Crosby Chapman, Western Reserve University, Cleveland, Ohio.
Reference: Chapman, J. C. "The measurement of physics information," *School Review*, 27:748-49, December, 1919.
The test with answers is reproduced. Results for 158 high-school pupils are tabulated.

Jones' Union Science Tests. These tests form an elaborate series covering many of the topics of the subject. There is a separate test for each topic.
Address and send results to: Franklin T. Jones, 10109 Wilbur Avenue, Cleveland Ohio.

Reference: Jones, F. T. "Practice exercises in physics," *School Review*, 26:341-48, May, 1918.

Describes the Union Science Tests—a series in physics. Test H_7-Test H_9, and Test H_8 are reproduced, with results analyzed.

Randall, Chapman and Sutton's Test in Mechanics. A simple test of a set of progressive problems in mechanics.

Address authors: Western Reserve University, Cleveland, Ohio.

Reference: Randall, D. P., Chapman, J. C. and Sutton, C. W. "The place of the numerical problem in high school physics," *School Review*, 26:39-43, January, 1918.

The test is reproduced. Results from experimental testing with four high schools are given.

Starch's Physics Test. This test is composed of 75 sentences which are arranged in the form of a completion test. The degree to which the pupils have mastered the work in physics covered by these sentences is shown by their ability to fill in the missing words, phrases, and numbers. The test is broken up into topics and standard scores for each topic are printed on the test sheet. There are 28 sentences on mechanics, eight on heat, nine on light, and 21 on magnetism and electricity.

Publisher: Daniel Starch, University of Wisconsin, Madison, Wisconsin. Also obtainable from Bureau of Educational Research, University of Illinois, Urbana, Illinois.

Reference: Starch, Daniel. *Educational measurements.* New York: Macmillan, 1916, chapter XIV.

XIV. Vocational Subjects and Manual Training

Leavitt's Preliminary Test for Manual Arts. As the title suggests this is only a preliminary test, not designed by the author for general use.

Address: F. M. Leavitt, Associate Superintendent in Charge of Vocational Education, Pittsburgh, Pennsylvania.

Reference: Leavitt, F. M. "Standardized measurements in the field of industrial arts," *Industrial Arts Magazine*, 8:132-8, April, 1919.

Brief discussion of methods of producing and using standardized tests is given. The preliminary scale for manual arts is reproduced.

Wardner's Test for Knowledge of Tools. This is a modification of one of the Army Trade Tests. Pupils are asked to identify 81 tools. Test is reproduced.

Address: C. A. Wardner, Director of Cooperative Trade Schools, Springfield, Vermont.

Reference: Wardner, C. A. "Applying the army trade tests to vocational schools," *Industrial Arts Magazine*, 8:402-03, October, 1919.

A reproduction and brief discussion of this test which is being used in the Cooperative Trade School, Springfield, Vermont.

(The End)

News Items and Communications

National Intelligence Tests

In the spring of 1919 the General Education Board appropriated $25,000 to the National Research Council to be used for the preparation of methods of measuring the intelligence of children in the elementary schools. A committee consisting of Dr. M. E. Haggerty, Dr. L. M. Terman, Dr. E. L. Thorndike, Dr. G. M. Whipple, and Dr. R. M. Yerkes, Chairman, was organized to undertake this task.

This committee has prepared two scales, the tests of which are listed below.

Tests of Scale A	Tests of Scale B
Arithmetical Reasoning	Computation
Sentence Completion	Information
Logical Selection	Vocabulary
Synonym-Antonym	Analogies
Symbol-Digit	Comparison

Either of these scales may be used alone, but the committee expects to recommend re-examination by the second scale after an interval of at least one day, in order that the reliability of the intelligence measurement may be increased. It is deemed preferable by the committee to use two short scales in the way suggested, rather than to use a single longer scale, the application of which would require approximately an hour of continuous work. Examination by either of the proposed scales, A or B, will require approximately thirty minutes.

Materials have been prepared by the committee for ten alternative forms of each of the ten tests which have been selected for use, but it is planned to complete for immediate publication only five forms of each test and of each scale. The remaining materials will be held subject to later development as needed.

It has been arranged that the World Book Company of Yonkers, New York, shall publish these new tests under the title "National Intelligence Tests." It is expected that the materials, which will consist of record blanks, examiner's guide, and scoring keys, will be ready for distribution in the spring of 1920. Royalties from the publications will be used for further research on methods of measuring intelligence and the practical application of such methods.

ROBERT M. YERKES
Chairman

Need to Teach Spelling of Derivatives

The conclusions of E. J. Ashbaugh concerning the necessity of teaching derived forms in spelling[1] have been confirmed by the results obtained in tests given by the Department of Educational Investigation and Measurement, Boston, Massachusetts. While his data are based on tests given in the lower elementary grades (II to V), the data of this department are based on results obtained in grades VI and VII. There is also this difference in the two investigations. The words tested by Ashbaugh may or may not have been studied by the pupils since the tests were not organized on that basis, while in the Boston tests the root words given were specifically chosen from the Boston minimum spelling list which the pupils had studied.

Undoubtedly the adding of a suffix to a word makes that word harder to spell and we have no right to presuppose that pupils can spell derived words the roots of which they are able to spell. This is shown con-

COMPARISON OF ROOTS AND DERIVATIVES

(All root words occur in sixth-minimum list)

Root Words	Percent Correct	Derivatives	Percent Correct
abbreviate	78	abbreviates	75
absence	85	absences	75
accidents	90	accidents	77
arrange	92	arranging	86
attack	94	attacked	87
		attacking	94
boundary	87	boundaries	67
century	93	centuries	77
commence	85	commenced	81
		commencing	74
conquer	89	conquered	80
deceive	80	deceiving	70
decide	87	decided	85
		deciding	76
describe	82	describing	76
industry	96	industries	79
judge	98	judges	93
material	84	materials	75
mosquito	94	mosquitoes	59
recognize	87	recognized	85
return	98	returned	93
		returning	95
salary	96	salaries	77
telegraph	98	telegraphed	90
telephone	98	telephoned	95
ventilate	87	ventilated	85
wharf	99	wharves	69

[1] Reported at February 1919 meeting of the National Association of Directors of Educational Research and in *Journal of Educational Psychology*, March, 1919, p. 143.

clusively in the results of spelling tests given in Boston in 1916,[1] 1917, and 1919. In each of the years 1916 and 1917 a seventh-grade division of approximately 1,000 pupils was given a list of root words to spell; another seventh-grade division was given derivatives of these same words; i.e., the first division spelled such words as "health," "accurate," and "promise," while the second division spelled "healthy," "accurately," and "promised." With few exceptions there was a falling off, and in some cases a very large falling off, in the percent of accuracy with which the derived forms were spelled, even in cases where the difference in spelling was only the addition of "s" or "ing" with no change of the existing root.

As a result of these findings the teachers were urged to teach in their spelling lessons the more common derivatives in connection with the root word.

In the spelling test of May 1919, 35 sixth-grade classes spelled the following list of 23 root words (all included in the sixth-grade minimum list) and 27 derivatives of those words with the percent of accuracy indicated in the table on page 322.

In every case but one ("attacking") the derivative proved itself a harder word to spell even though the teachers had been urged to teach the common derivatives. Since it is as important that the pupils know how to spell such common words as "judges," "mosquitoes," etc., as well as the root words, *every common derived form must be considered as presenting a new spelling problem and must be taught as such.*

HARRIET M. BARTHELMESS

Conference on Educational Measurements

The Seventh Annual Conference on Educational Measurements will be held at Indiana University, Friday and Saturday, April 23 and 24. Indiana University was the first institution to hold an annual conference on educational measurements. These meetings are well attended by the school men of the state; and the university has brought to this meeting year after year speakers of national reputation. The committee in charge of the conference this year has issued the following preliminary announcement:

"The principal speakers for the Seventh Annual Conference on Educational Measurements will be Dr. Edward L. Thorndike of Teachers College, Columbia University, and Professor C. E. Seashore of the University of Iowa. Dr. Thorndike's discussions will be along the general lines of intelligence examinations for college entrance, standardization of textbooks, and the dangers of indiscriminate use and misuse of educational tests. Professor Seashore, who has done some remarkable work in the particular field he will present, will bring to the Conference the results of his research in the measuring of special abilities, particularly the measuring of ability in music."

[1] See Bulletin XI of the Department of Educational Investigation and Measurement, Boston, Massachusetts.

Southington-Plymouth English Vocabulary Scale

A number of different types of tests have been devised for measuring the silent reading vocabulary of pupils. The best known of these probably is the Thorndike's Visual Vocabulary Scales A2 and B. In these the pupil is asked to indicate the class to which the word belongs as flower, animal, or boy's name. In other vocabulary tests, the pupil is asked to check the words he knows the meaning of and to write out the meaning of those concerning which he is uncertain. Both of these plans have obvious limitations. The Southington-Plymouth English Vocabulary Scale devised by Ernest C. Witham makes use of a different device. Fifty English words are given. On the opposite page their brief definitions as given by Webster's *International Dictionary* are printed. The pupil is to record before each definition the number of the word which it fits.

Most of the words were selected from Ayres Spelling Scale. A few additional words were taken from Thorndike's Visual Vocabulary Scale A2, series *x*. The test has been tentatively standardized for the eighth grade and the high school. Copies of the test can be secured by addressing Superintendent Ernest C. Witham, Southington, Connecticut.

Classifying Pupils in English

Mr. George N. Cade, Supervisor of Elementary Grades, Springfield, Illinois, gave Monroe's Standardized Silent Reading Test II to the VIIIA pupils who were promoted to high school at the end of the first semester. The principal of the high school has classified these pupils in the English department on the basis of their ability to read silently as shown by this test. Since ability to read silently is fundamental to practically all of the work of the high school, classification of pupils on the basis of this ability may be expected to effect a helpful organization.

A Non-Verbal Intelligence Test

The Myers Mental Measure by Caroline E. Myers and Captain Garry C. Myers, claimed by the authors to be the first group intelligence test applicable to all ages, is recommended for use in every grade since the directions are given orally and since the exercises consist wholly of pictures. It can be given to illiterates as well as to those who can read. The time for giving it is not more than twenty minutes. The publisher is The Sentinel, Carlisle, Pennsylvania.

The Virginia School Survey Tests

Professor M. E. Haggerty of the University of Minnesota devised for use in the Virginia School Survey two tests of general intelligence, Delta 1 for grades one to three and Delta 2 for grades three to nine; and a test for measuring achievement in reading, Sigma 1 for grades one to three. These tests have recently been published and made available for general use by the World Book Company of Yonkers, New York. The tests are accompanied by a manual of directions and keys for scoring. Their use in the Virginia Survey and in a number of cities elsewhere in the country insures their standardization.

National Association of Directors of Educational Research

(E. J. Ashbaugh, *Secretary and Editor*)

Even though it is several weeks since the meeting of the Department of Superintendence at Cleveland, this is the first opportunity to tell you, who were not fortunate enough to attend, something of the meetings of our association at that time.

The first annual closed meeting on Monday was a decided success. Although conflict of time with the meetings of the College Teachers of Education prevented a few members from attending, the number present was quite satisfactory and the program even more so. Though announced as a closed meeting, nothing short of a corporal's guard would have kept some interested non-members out. It was more than curiosity on their part, since once inside they remained throughout the meeting.

The forenoon was given over to the report of the Committee on Standards (Mr. Courtis, Chairman) and a paper on "Diagnosis" by Mr. Kallom of Boston, the paper being presented by Miss Barthelmess in his absence. The report of the Committee on Standards attempted to accomplish its task by: (1) pointing out the problems to be solved, the ends to be achieved; (2) indicating so far as is at present possible the criteria to be used in judging of values; (3) providing an incentive to induce workers to apply these criteria; and (4) leaving it to a process of natural selection to determine which method shall survive. The association plans to make available to all its members as well as to the public the terms, methods, and practices approved. The association passed a motion for the reorganization of the committee with provision for subcommittees in a number of special fields.

The paper on "Diagnosis" set forth the method the Boston department is using in acquainting the assistant superintendents and masters with the position of classes relative to city standards. Although, as a result of these reports, considerable effectiveness is being registered in the teaching; it was admitted by the speaker that as yet the department has been unable on account of lack of resources to accompany its reports to the teachers with suggestions as to remedial measures.

At 12:30 luncheon was served in "our" private dining room. Thirty of the members and their guests participated.

The afternoon program included two papers on "The Evaluation of Tests"; one by Dr. Monroe and the other by Dr. T. L. Kelley. Both speakers led the members into a very deep discussion of the need for the statistical methods involved in the determination of the validity and reliability of tests. It

involved a consideration of technic impossible for the majority of us to practice but represented a phase of the testing work in which far too little has been done.

President Buckingham gave the members a brief history of the evolution of our official organ, the JOURNAL OF EDUCATIONAL RESEARCH. We were delighted to learn that it is exceeding the hopes of its founders, that it is already on a self-supporting basis, and that the first issue has to be reprinted to supply the demand for it.

At six o'clock we again broke bread together in the comfortable adjoining room which the hotel management had assigned to us. By this time we felt as if we owned it.

On Thursday a program consisting of eleven numbers was presented. Judging from the attendance and the comment of those who heard our offerings, the program was a success. Most of the papers will be published in the JOURNAL OF EDUCATIONAL RESEARCH.

President Brown of Oshkosh led off with his paper on "Formulation of Method in Reading in the Light of Recent Investigations." As the chairman said in his introduction, Mr. Brown is a conspicuous example of a school man who has become an executive without ceasing to participate in research work. The morning program was organized about the topic of reading, and Mr. Brown's paper afforded a good introduction to the work of the morning.

Dr. Theisen handled the subject of "Provisions for Individual Differences in the Teaching of Reading" in an effective manner. He has gathered material from more than one hundred teachers on the provisions which they are making and the provisions which they feel ought to be made for adjusting the teaching of reading to differences in ability among children. The report was interesting because it showed just what the teachers are thinking about on the important matter of fitting instruction to the needs of pupils.

Mr. Kallom was to have presented a paper on "Reproduction as a Measure of Ability." He was, however, at the last minute, obliged to remain at home on account of illness in his family. We were sorry indeed not to greet Kallom but his paper was so capably presented by Miss Barthelmess, his assistant at the Boston Department of Research, that we almost forgot our disappointment. The point of the paper was that children reproduce what they have read in terms of their experience. Some of the responses of children were interesting not only from a scientific point of view but also were worthy of a place among the "bright sayings" of children.

Doctor—or shall we say Colonel—Ayres dodged the use of slides and brought his own charts. This was a wise precaution on his part, for we were having trouble with our lantern. These charts he comfortably set up and proceeded to offer an admirable paper on "The Comparative Effectiveness of State School Systems." He began by going back to 1750 at which time an Italian article was published on the high cost of living. In this article index numbers were used practically as they are now employed in economics. The speaker discussed the meaning and use of the index number and showed how it could be applied in education. He made the application on eight items to the schools of each of the states. These states were ranked on the basis of

each of the eight items according to their index numbers. Comparisons were also made with Dr. Ayres' previous study of conditions in 1910.

Last year, our association authorized a committee on standards. The president, however, seemed to have been particularly successful in selecting men who were about to retire from research work and from membership in the organization. The committee had no opportunity, therefore, to work out a real committee report. Mr. Courtis, however, as chairman, determined to have a report, so he wrote one himself. This he presented in part as the first number on the afternoon program. While certain recommendations were made for the guidance of the association and through it for the guidance of research workers in general, the paper consisted largely of the presentation of evidence to show that we really do not know exactly what our tests measure. As an indication of the analysis to which our testing instruments must be subjected, the paper was important.

Dr. Thorndike was to have presented a paper on "Intelligence Tests for College Entrants." He was, however, unable to attend—and again we have the "flu" to thank for his inability to respond. Dr. Thorndike sent his paper but, owing to the shortened program consequent upon the delay in getting started, the paper was not read. It will be published in an early number of the JOURNAL OF EDUCATIONAL RESEARCH.

Mr. Dickson—Dickson all the way from the Coast—gave a report on the use of intelligence tests in the classification of eighth and ninth-grade pupils. He demonstrated the reliability of these tests as administrative devices for the homogeneous grouping of pupils for educational guidance, and for providing specialized instruction for superior children.

Mrs. Fleming of the Wisconsin State Department presented her paper on "Reading Ability and Intelligence." Mrs. Fleming had results on the use of the Thorndike reading test in Wisconsin and Virginia. From the Wisconsin results she is tempted to conclude that there is a close correlation between reading ability and intelligence but the Virginia results seemed contrary. While not entirely committing herself, she seemed to feel that such correlation did exist wherever regular school attendance and emphasis upon reading were generally operative.

Dean Kelly presented a study of "Three Methods of Teaching the Fundamentals of Arithmetic." The methods were by use of (1) Courtis Practice Tests; (2) Studebaker Practice Tests; (3) Usual teacher-devised drill material. The results were definitely favorable to scientifically devised practice material, the "Courtis" getting, in general, the greatest gains for classes but the "Studebaker" evidencing greater corrective effect.

Mr. Burgess, one of our new members in the Russell Sage Foundation, presented "A Trend Analysis of Teachers' Salaries." By a study of the wages of skilled and unskilled laborers and of rural and city teachers during the past century and by converting the wages into index numbers, he showed the following facts: (1) wages of one kind follow the trend of wages of another kind much more closely than they do the cost of living, (2) wages of skilled and unskilled labor respond more quickly in economic crises than do teachers' salaries, (3) rural teachers' salaries and wages of unskilled labor tend to be

nearly equal except in times of crisis, (4) city teachers' wages tend to follow the same general curves as those of rural teachers and of skilled and unskilled laborers; but they lie between the wages of the last named groups, (5) times of high prices show a decided drop in the index of teachers' salaries.

Mr. Richardson showed some results attained by "The Campaign Method in Elementary Education" with special reference to arithmetic work in the four fundamental operations with whole numbers. The method consists of taking individual results directly to each child and interesting him in making a record for himself. Practice examples were furnished the pupils. Although the campaign was carried on for a few weeks only, results showed large and permanent gains.

The Fifth Annual Banquet and the evening program were the climax of the week to the members of our association. More than forty members and guests were present. Serious thought and kindly humor characterized the evening to the delight of all.

Following the practice of the past, the secretary presented one name for election to honorary membership. The association honored itself this year by electing Dean Wm. F. Russell of the College of Education, State University of Iowa. Dean Russell responded to the election with a striking allegory, portraying the evolution of the school superintendent in terms of the chauffeur.

Mr. Packer presented very briefly the need of a definite administration of supervisory work through a department of research. Such a plan is now in operation at Detroit. By this means, he said, much supervision which is ordinarily wasted, may be applied where it is most needed. His plan contemplates the fuller recognition of the building principal as a supervisor.

Dr. Rugg spoke on the necessity of scientificially devised text books as the solution of much of the present school inefficiency—a solution which can be applied even before we have an adequately trained body of teachers.

The first annual presidential address was given, summarizing the progress of the past year in the fields of particular interest to our membership. The magnitude of this progress was surprising to most of us.

At the business meeting, the following officers were elected for 1920-1921: President, M. E. Haggerty, Director of Bureau of Co-operative Research, University of Minnesota; Vice-president, Virgil E. Dickson, Director of Research and Guidance, Oakland and Berkeley, California; Secretary-Treasurer, E. J. Ashbaugh, Director of Bureau of Educational Service, State University of Iowa, Iowa City, Iowa.

Much of the business so far as discussion was concerned had been conducted at the closed session on Monday. Accordingly matters were disposed of quickly. In addition to the election of officers, the president was authorized to appoint a committee with power to act in the matter of affiliation with the National Research Council. He was also requested to appoint a permanent Standardization Committee and a committee to decide the day of the convention week on which our closed session shall hereafter be held. Dues were raised from $2.00 to $3.00 beginning with 1921.